UNIVERSITY OF SOUTHAMPTON

Variability and uncertainty in measuring sea surface temperature

by

Werenfrid Wimmer

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Natural and Environmental Sciences
Ocean and Earth Sciences

December 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

OCEAN AND EARTH SCIENCES

Doctor of Philosophy

Variability and uncertainty in measuring sea surface temperature

by Werenfrid Wimmer

Sea Surface Temperature (*SST*) measurement is one of the most easily obtainable climate variables. However, it is challenging to meet the required absolute accuracy and long term stability whether the data are derived by *in situ* or satellite measurements. This study explores the quality of *SST* measurements, in particular those derived by the Advanced Along Track Scanning Radiometer (AATSR) and *in situ* measurements recorded by the shipborne Infrared Sea surface temperature Autonomous Radiometer (ISAR), which are used for validating AATSR data. Its broad objective is to improve understanding of measurement uncertainties in order to quantify the quality of satellite derived *SST* used for climate records.

The uncertainties of *in situ* measurement by ISAR have been analysed and modelled in order to estimate an independent measurement uncertainty for every *SST* data point in the ISAR records. In a complementary study the separate uncertainties of the *SST* as observed by AATSR, ISAR and ship-based hull-mounted thermometry (SST_{depth}), when observing the same track, have been resolved by means of three way uncertainty analysis. This not only serves to verify the ISAR uncertainty model but also demonstrates the effectiveness of using shipborne radiometry in preference to in water thermometry from ships or buoys for validating satellite *SST* products. A third area of study concerns the errors and uncertainties when comparing satellite and *in situ* observations, which result from failure to properly match the *in situ* observations to what the satellite "sees". A new method has been developed for classifying the "$match-up$ quality" of each data pair. Its use is demonstrated to show that the quality of AATSR data may be better than classical validation $match-ups$ have shown. Finally, consideration is given to how these new methods for estimating measurement uncertainties may increase the confidence in *SST* climate datasets.

Contents

Abstract iii

Table of Contents v

List of Figures ix

List of Tables xvii

Declaration of Authorship xix

Acknowledgements xxi

Acronyms xxv

Symbols xxix

1 Introduction 1
- 1.1 Preface . . . 1
- 1.2 Thesis structure . . . 2
- 1.3 The relevance and importance of SST . . . 4
- 1.4 Aims and Objectives . . . 5

2 Background 7
- 2.1 Temperature in the upper ocean . . . 8
- 2.2 The skin layer and the cool skin effect . . . 11
- 2.3 Radiometric temperature measurements . . . 14
- 2.4 Diurnal variability . . . 18
- 2.5 Climate relevance of temperature measurements . . . 18
- 2.6 Operational requirements . . . 21
- 2.7 $In-situ$ Radiometers . . . 22
  - 2.7.1 M-AERI . . . 22
  - 2.7.2 SISTeR . . . 24
  - 2.7.3 CIRIMS . . . 25
- 2.8 Measurements on Ships of Opportunity and Voluntary Observing Ships . . . 26
- 2.9 Study area . . . 27

3 Measurements 29
- 3.1 Satellite SST_{skin} . . . 30
- 3.2 *In situ* radiometric measurements - ISAR . . . 32
  - 3.2.1 Basic design of ISAR . . . 32
  - 3.2.2 A self calibrating radiometer . . . 35

3.2.3 Radiometric measurements made by ISAR . . . 37
3.2.4 Revised radiometric calculation . . . 41
3.2.5 Quality control of radiometric measurements . . . 44
3.2.5.1 ISAR engineering data . . . 44
3.2.5.2 Verification against a black body source . . . 58
3.2.5.3 Traceability to international standards . . . 59
3.3 Standard ISAR configuration . . . 64
3.4 Installation on the Pride of Bilbao . . . 64
3.4.1 Sea Surface Temperature at depth . . . 65
3.4.1.1 SBE 48 . . . 65
3.4.1.2 Hull Thermistors . . . 66
3.4.1.3 CPR temperature probe . . . 66
3.4.2 Ancillary measurements . . . 67
3.5 Installation on the Cap Finistere . . . 70
3.6 Overview of ISAR data . . . 73
3.7 AATSR - ISAR validation data . . . 76

4 ISAR uncertainties 79
4.1 Introduction . . . 79
4.2 An uncertainty budget . . . 80
4.2.1 Uncertainty propagation . . . 81
4.2.2 Type A uncertainties . . . 83
4.2.3 Type B uncertainties . . . 83
4.2.3.1 Sea water emissivity . . . 83
4.2.3.2 Internal black body emissivity . . . 86
4.2.3.3 Thermistors . . . 87
4.2.3.4 Resistance to Temperature approximation . . . 88
4.2.3.5 KT15 . . . 88
4.3 Estimating the ISAR uncertainty . . . 91
4.4 Summary . . . 96

5 Uncertainty analysis of *SST* measurements 97
5.1 Introduction . . . 97
5.2 Three way uncertainty analysis . . . 98
5.2.1 The OC method . . . 98
5.2.2 The TCM method . . . 100
5.3 A model for estimating the uncertainty . . . 101
5.3.1 A model for calibration . . . 102
5.3.2 Indeterminacy . . . 103
5.3.3 Variances of estimators . . . 104
5.3.4 A practical problem . . . 104
5.4 Data . . . 105
5.4.1 Data quality and correlation . . . 107
5.5 Results . . . 110
5.6 Discussion . . . 111

6 Uncertainty analysis of the *match − up* process 113
6.1 Introduction . . . 113
6.2 Classic Match-ups . . . 114
6.3 The *match − up* uncertainty budget . . . 117
6.4 The estimation of the *match − up* uncertainty . . . 121

6.5 Design of a quality value . . . 124
6.5.1 Estimating the QI thresholds . . . 125
6.6 Results . . . 132
6.7 Discussion of the results . . . 145
6.8 Implications of the quality indicator (QI) . . . 147

7 Conclusion **149**
7.1 Summary and Discussion of the main achievements . . . 149
7.1.1 ISAR uncertainty . . . 149
7.1.2 Uncertainty analysis of *SST* measurements . . . 150
7.1.3 Uncertainty analysis of the $match - up$ process . . . 151
7.2 Scientific relevance . . . 152
7.3 Future work . . . 154

Appendix A Further three way uncertainty plots **157**

Appendix B Match-up QI results **163**
B.1 Statistics . . . 164
B.2 Plots . . . 172
B.2.1 AATSR N1 data with normal statistics . . . 172
B.2.2 AATSR confidence value (CV) 3 data with normal statistics . . . 181
B.2.3 AATSR CV 4 data with normal statistics . . . 184
B.2.4 AATSR CV 5 data with normal statistics . . . 189
B.2.5 AATSR N1 data with robust statistics . . . 198
B.2.6 AATSR CV 3 data with robust statistics . . . 207
B.2.7 AATSR CV 4 data with robust statistics . . . 210

Appendix C SSES for AATSR **215**

Appendix D Validation paper **217**

References **233**

List of Figures

2.1 Schematic view of the upper ocean temperature profile representing a) night-time and day time high wind conditions in blue and b) day time low wind conditions in red. Note that the temperature differences as given are typical but in reality can vary from these values quite strongly. Values of (SST_{skin} - $SST_{foundation}$) of up to 6 K have been recorded. Also note that the SST_{depth} corresponds to a SST measurement made by a thermometer in the water column, which could be at any depth from 0 depth to the bottom of the sea, although for this study SST_{depth} refers to as a temperature somewhere between $SST_{subskin}$ and $SST_{foundation}$. After Donlon *et al.* (2007) 9

2.2 Behaviour of the oceanic skin layer under three primary physical regimes: T_A, T_S, and T_B represent, respectively, the air temperature, oceanic skin temperature, and bulk ocean temperature; Q_E, Q_H, and Q_N represent the latent, sensible, and net radiative heat fluxes; and τ_a and τ_w represent the surface stress in the air and the water (from Emery *et al.*, 2001). . . . 12

2.3 Vertical temperature profiles between 0.1 m above(atmosphere) and 0.1 m below (ocean) computed by the INTRA model Eifler and Donlon (2001). Triangles represent observed mean Sea Surface Temperature at the Skin interface (SST_{skin}); squares represent maximum SST_{skin} and diamonds represent Sea Surface Temperature at depth (SST_{depth}). 14

2.4 Black body radiation. The blue line represents the thermal emission for a temperature of 100 K, the green line 270 K, the yellow line 320 K, the red line 1000 K and the black line is approximately the temperature of the sun of 5500 K. The two shaded areas show in yellow the visible spectrum and in red the infra-red spectrum between 3.7 and 12 μm in the main plot. The zoomed out section in the top right corner of the plot show again 270 K and 320 K as green and yellow lines. The shaded areas are four window regions used for SST measurements from satellites. Shaded grey is 3.7 μm , magenta is 7.7 μm , blue 11 μm and red 12 μm 16

2.5 Sea Surface Emissivity (ε) (labelled as SSE) dependence on wavelength and observation angle, as given by the model of Wu and Smith (1997) for a wind speed $U = 0ms^{-1}$, from Niclos *et al.* (2009) 17

2.6 The Miami 2001 inter-comparison experiment. On the far right the Marine Atmospheric Emitted Radiance Interferometer (M-AERI) is visible, next is the Calibrated InfraRed In situ Measurement System (CIRIMS), to the left of CIRIMS is ISAR 01 and far left is Scanning Infrared Sea Surface Temperature Radiometer (SISTeR). From Barton *et al.* (2004) . . 23

2.7 The M-AERI on the Explorer of the Seas. From Minnett (2004). 23

2.8 The SISTeR. From Nightingale (2007) 24

2.9 The CIRIMS installed on the R/V Ron Brown. From Jessup (2003). . . . 26

2.10 Overview of the study area with a typical ship track super-imposed on the bathymetry of the area 27

3.1 AATSR view geometry schematics, from Llewellyn-Jones *et al.* (2001) . . 31
3.2 (a) ISAR optical path showing the main components of the ISAR optical system; the instrument detector (Heitronics KT15.85D (KT15)), ZnSe plane Window, scan drum and gold mirror, protective bush (no longer used) and scan drum aperture and calibration black body (BB). (b) location of the ISAR calibration BB cavities in the main instrument body showing the main views made by the ISAR radiometer; sea, sky, BB1, BB2. From Donlon *et al.* (2008) 33
3.3 External view of the ISAR instrument showing the shutter open. The optical rain sensor and the Global Positioning System (GPS) antennae is pictured in front of the instrument. 34
3.4 Geometrical quantities and radiative components that must be considered when measuring the SST_{skin} temperature of the ocean surface. The figure shows a dual port radiometer mounted aboard a ship at a height h above the sea surface viewing the sea surface at a zenith angle θ. From Donlon *et al.* (2008). 38
3.5 The ISAR engineering plot for D22, the image shows the ship track in the top panel and scan drum position (in blue, left hand scale) with the shutter state (red, right hand scale) in the bottom panel.The green diamonds show that the shutter has reached its end position. 48
3.6 The ISAR engineering plot for D22, the image shows the instrument thermistor temperatures; in the top panel various thermistors that are distributed throughout the instrument are shown, in the middle panel the ambient temperature black body thermistors temperatures are shown and in the bottom panel the heated temperature black body thermistors temperatures are shown 49
3.7 The ISAR engineering plot for D22, the image shows the instrument thermistor voltages; in the top panel various thermistors which are distributed throughout the instrument are shown, in the middle panel the ambient temperature black body thermistors voltages are shown and in the bottom panel the heated temperature black body thermistors voltages are shown 50
3.8 The ISAR engineering plot for D22, showing the detectors response in the top panel as voltages and in the bottom panel as brightness temperatures calculated with an emissivity of 1. 51
3.9 The ISAR engineering plot for D22, shows the GPS module extra outputs - speed over ground (SOG), course made good (CMG), Magnetic variation (VAR) - and the PNI roll and pitch sensor (PNI) module output in the bottom three panels. 52
3.10 The ISAR engineering plot for D22, shows the input power voltage to the instrument in the top panel, the reference voltage used for the black body thermistors in the second panel, the SST in the third panel, the shutter state in the fourth panel and the rain gauge signal in the fifth panel . . . 53
3.11 The ISAR engineering plot for D22, shows a zoomed section of the rain gauge signal (blue) with the shutter state superimposed in red. 54

3.12 The ISAR engineering plot for D22, shows the sky brightness temperature in the top panel and the sea brightness temperature (in blue) and the SST_{skin} (in magenta) in the bottom panel. 55
3.13 The ISAR engineering plot for D22, shows the gain factor Equation 3.20 calculated with temperatures in the top panel and radiances in the bottom panel. The blue line is a 100 point moving window average. 56
3.14 The ISAR engineering plot for D22, shows the window temperature split with viewing angle in the top panel, the temperature difference between the ambient and heated black body as temperature difference in the second panel, as voltage difference in the third panel and as radiance difference in the bottom panel. The blue line in the bottom three panels is 100 point moving window average. 57
3.15 The Combined Action for the Study of the Ocean Thermal Skin (CASOTS) II black body with the calibration cradle for the ISAR. 59
3.16 Calibration plot of ISAR 03 before Deployment 22 61
3.17 Calibration plot of ISAR 03 after Deployment 22 62
3.18 Calibration plot of ISAR 03 after Deployment 22 with the revised SST (see section 3.2.4) calculation applied 63
3.19 The instrument installation on the Pride of Bilbao (PoB) 68
3.20 ISAR mounted on the starboard bridge roof on the PoB. 69
3.21 Close up of the ISAR mounted on the starboard bridge roof on the PoB. . 69
3.22 Cap Finistere with the ISAR location labelled. 71
3.23 ISAR mounted on the starboard bridge walkway on the Cap Finistere (CpF) 72
3.24 Close up of the ISAR close up on the CpF 72
3.25 ISAR data collected from 2004 to 2011 73

4.1 Flow chart of the ISAR SST processor. R_{sky} follows the same path as R_{sea}. Boxes coloured in blue show Type A uncertainties, boxes coloured in red show Type B uncertainties and boxes in red and blue show that the particular box has both Type A and Type B uncertainties. 82
4.2 Plot of the effect the uncertainty in the seawater emissivity has on the SST processor uncertainty. Here BT_{sea} is the simulated Brightness Temperature of the sea view, BT_{sky} is the simulated Brightness Temperature of the sky view, SST the calculated Sea Surface Temperature and std_{SST} the resulting uncertainty of the SST estimation. 85
4.3 Plot of the change in uncertainty depending on BT_{sky} and SST. Where BT_{sky} is the simulated Brightness Temperature of the sky view, SST the calculated Sea Surface Temperature and std_{SST} the resulting uncertainty of the SST estimation. 86
4.4 Section through the an ISAR calibration BB radiance cavity showing the re-entrant cone design, thermal shroud and location of thermistors used to determine the radiative temperature of the BB. The inner surfaces of the BB are coated with Nextel velvet black 811-21 paint. The emissivity of this design is 0.9993. Image from Donlon *et al.* (2008) 87
4.5 The YSI 46041 Thermistors resistance to temperature data is shown in the top panel; the middle panel shows the residuals of the residuals of the Steinhart-Hart approximation in red and a third order polynomial fit in blue and teh bottom panel shows a tenth order fit of the residuals shown in the middle panel. 89

4.6 The KT15.85D detector response shown as relative voltage vs target temperature. . . . 90
4.7 The KT15.85D detector filter response as used in ISAR. The numbers in the label box of the plot refer to the KT15.85D serial number as used in the different ISAR. . . . 90
4.8 ISAR uncertainties for the data collected between 15.07.2011 and 20.07.2011 on the CpF. The top panel shows the measurement uncertainty. the bottom left panel shows the uncertainty split into fractions of Type A and Type B uncertainty and the bottom right panel shows the fraction of instrument and measurement uncertainty. . . . 92
4.9 ISAR uncertainties for the data collected between 15.07.2011 and 20.07.2011 on the CpF. Panel a shows the total uncertainty, panel b the uncertainty in over the $BT_{sea} - T_{BBamb}$ difference, panel c shows the uncertainty plotted against the latitude and panel d a histogram of the uncertainty. Panel e shows the uncertainty plotted against the Brightness Temperature of the sea view (BT_{sea}), panel f shows the uncertainty plotted against the Brightness Temperature of the sky view (BT_{sky}), panel g shows the uncertainty plotted against the temperature of the ambient BB and panel h shows the uncertainty plotted against the SST_{skin}. The blue dots represent data where BT_{sea} is colder than the ambient BB and red dots show data where BT_{sea} is between the ambient BB and the hot BB. . . . 94
4.10 ISAR uncertainties for the data collected between 15.07.2011 and 20.07.2011 on the CpF plotted along the ship track. . . . 95

5.1 Histogram of the data difference for AATSR D3 data. . . . 108
5.2 QQ-plot of the data difference for AATSR D3 data. . . . 108
5.3 QQ-plot of the data difference for AATSR D3 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot. . . . 109

6.1 Examples of $match - up$ situations encountered in the construction of a $match - up$ database. (a) Point sample when there is no cloud. Match the *in situ* sample closest to the time of the overpass to the pixel in which it lies. (b) Point sample obscured by cloud. Match the *in situ* sample closest to the time of the overpass to the closest cloud-free pixel. The search radius needs to be limited to N pixels. (c) Along-track sensor such as ISAR in cloud free conditions. Match the *in situ* sample closest to the time of the overpass to the pixel in which it lies. (d) Along-track sensor in cloudy conditions. From Wimmer and Robinson (2008a) . . . 116
6.2 AATSR image from 16.03. 2009, showing a $match - up$ scene where $match - up$ windows work well. The AATSR pin shows the AATSR pixel matched to ISAR with a grade 2b $match - up$ window. From Hopkins (2010). . . . 117
6.3 AATSR image from 12.04.2009, showing a $match - up$ scene where $match - up$ windows give a high difference between the $match - up$ pair. The AATSR pin shows the AATSR pixel matched to ISAR with a grade 2b $match - up$ window. From Hopkins (2010). . . . 118
6.4 Schematics of the de-trending process. Shown in blue is the SST data as captured by AATSR, in black is the fitted flat plane and in red the de-trended SST data. The green pixel shows the $match - up$ pixel and grey pixels show cloudy data. . . . 122

6.5 Quality indicator plot showing how the different parameters set the quality level for the $match - up$ pair for the AATSR dual view night time $match - up$ from the training data set. The the colour of the dots represents in all but the wind speed plot the QI, with dark green representing QI 5, light green QI 4, yellow QI 3 and red QI 0. 127
6.6 Quality indicator plot showing how the different parameters set the quality level for the $match-up$ pair for the AATSR night time $match-up$ from the training data set. Note that the bar plots were stacked on top of each other for better visibility and only show which indicator has the most impact in the given scene. 128
6.7 Quality indicator plot showing how the different parameters set the quality level for the $match - up$ pair for the AATSR dual view day time $match - up$ from the training data set. The the colour of the dots represents in all but the wind speed plot the QI, with dark green representing QI 5, light green QI 4, yellow QI 3 and red QI 0. 129
6.8 Quality indicator plot showing how the different parameters set the quality level for the $match-up$ pair for the AATSR day time $match-up$ from the training data set. Note that the bar plots were stacked on top of each other for better visibility and only show which indicator has the most impact in the given scene. 130
6.9 Robust Huber T distribution. 132
6.10 QI histograms for CV 5 and robust statistics for night time data 136
6.11 QI histograms for CV 5 and robust statistics for day time data 137
6.12 QI scatter plot for CV 5 and robust statistics for dual view night time data 138
6.13 QI scatter plot for CV 5 and robust statistics for dual view day time data 139
6.14 QI scatter plot for CV 5 and robust statistics for nadir view night time data 140
6.15 QI scatter plot for CV 5 and robust statistics for nadir view day time data 141
6.16 QI location plot for CV 5 and robust statistics for for all data 142
6.17 QI time and distance plot for CV 5 and robust statistics for for night time data 143
6.18 QI time and distance plot for CV 5 and robust statistics for for day time data 144
6.19 AATSR image from 24.09.2009, showing a $match - up$ scene with cloud contamination issues.. From Hopkins (2010). 146

A.1 Histogram of the data difference for AATSR D2 data. The coloured dots show a Gaussian distribution with the same mean and standard deviation as the data. 158
A.2 QQ-plot of the data difference for AATSR D2 data. 158
A.3 QQ-plot of the data difference for AATSR D2 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot. 159
A.4 Histogram of the data difference for AATSR N3 data. The coloured dots show a Gaussian distribution with the same mean and standard deviation as the data. 159
A.5 QQ-plot of the data difference for AATSR N3 data. 160
A.6 QQ-plot of the data difference for AATSR N3 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot. 160

A.7 Histogram of the data difference for AATSR N2 data. The coloured dots show a Gaussian distribution with the same mean and standard deviation as the data. . . . 161
A.8 QQ-plot of the data difference for AATSR N2 data. . . . 161
A.9 QQ-plot of the data difference for AATSR N2 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot. . . . 162

B.1 QI histograms for N1 and normal statistics for night time data . . . 172
B.2 QI histograms for N1 and normal statistics for day time data . . . 173
B.3 QI scatter plot for N1 and normal statistics for dual view night time data 174
B.4 QI results for CV 5 . . . 175
B.5 QI scatter plot for N1 and normal statistics for nadir view night time data 176
B.6 QI scatter plot for N1 and normal statistics for nadir view day time data 177
B.7 QI location plot for N1 and normal statistics for for all data . . . 178
B.8 QI time and distance plot for N1 and normal statistics for for night time data . . . 179
B.9 QI time and distance plot for N1 and normal statistics for for day time data . . . 180
B.10 QI histograms for ghrsst3 and normal statistics for day time data . . . 181
B.11 QI scatter plot for ghrsst3 and normal statistics for dual view day time data . . . 181
B.12 QI scatter plot for ghrsst3 and normal statistics for nadir view day time data . . . 182
B.13 QI location plot for ghrsst3 and normal statistics for for all data . . . 183
B.14 QI time and distance plot for CV 3 and normal statistics for for day time data . . . 183
B.15 QI histograms for CV 4 and normal statistics for night time data . . . 184
B.16 QI scatter plot for CV 4 and normal statistics for dual view night time data 185
B.17 QI scatter plot for CV 4 and normal statistics for nadir view night time data . . . 186
B.18 QI location plot for CV 4 and normal statistics for for all data . . . 187
B.19 QI time and distance plot for CV 4 and normal statistics for for night time data . . . 188
B.20 QI histograms for CV 5 and normal statistics for night time data . . . 189
B.21 QI histograms for CV 5 and normal statistics for day time data . . . 190
B.22 QI scatter plot for CV 5 and normal statistics for dual view night time data 191
B.23 QI scatter plot for CV 5 and normal statistics for dual view day time data 192
B.24 QI scatter plot for CV 5 and normal statistics for nadir view night time data . . . 193
B.25 QI scatter plot for CV 5 and normal statistics for nadir view day time data 194
B.26 QI location plot for CV 5 and normal statistics for for all data . . . 195
B.27 QI time and distance plot for CV 5 and normal statistics for for night time data . . . 196
B.28 QI time and distance plot for CV 5 and normal statistics for for day time data . . . 197
B.29 QI histograms for N1 and robust statistics for night time data . . . 198
B.30 QI histograms for N1 and robust statistics for day time data . . . 199
B.31 QI scatter plot for N1 and robust statistics for dual view night time data 200
B.32 QI results for CV 5 . . . 201

B.33 QI scatter plot for N1 and robust statistics for nadir view night time data 202
B.34 QI scatter plot for N1 and robust statistics for nadir view day time data . 203
B.35 QI location plot for N1 and robust statistics for for all data 204
B.36 QI time and distance plot for N1 and robust statistics for for night time data 205
B.37 QI time and distance plot for N1 and robust statistics for for day time data 206
B.38 QI histograms for CV 3 and robust statistics for day time data 207
B.39 QI scatter plot for CV 3 and robust statistics for dual view day time data 207
B.40 QI scatter plot for CV 3 and robust statistics for nadir view day time data 208
B.41 QI location plot for CV 3 and robust statistics for for all data 209
B.42 QI time and distance plot for CV 3 and robust statistics for for day time data 209
B.43 QI histograms for CV 4 and robust statistics for night time data 210
B.44 QI scatter plot for CV 4 and robust statistics for dual view night time data 211
B.45 QI scatter plot for CV 4 and robust statistics for nadir view night time data 212
B.46 QI location plot for CV 4 and robust statistics for for all data 213
B.47 QI time and distance plot for CV 4 and robust statistics for for night time data 214

List of Tables

2.1 Definitions of sea surface temperature proposed by Group for High Resolution Sea Surface Temperature (GHRSST) (Donlon *et al.*, 2007; Kawai and Wada, 2007) . . . 11
2.2 List of the Essential Climate Variable (ECV) as defined by Global Climate Observing System (GCOS) . . . 19

3.1 AATSR channels, from Llewellyn-Jones *et al.* (2001). . . . 30
3.2 List of data products for the level 2 AATSR . . . 32
3.3 ISAR traceability results 2009 (Theocharous *et al.*, 2010). Positive values mean ISAR is reading higher temperatures than the standard. . . . 60
3.4 CASOTS II traceability results 2009 (Theocharous and Fox, 2010). Negative values mean Combined Action for the Study of the Ocean Thermal Skin second generation black body (CASOTS II) is reading lower temperatures than the standard. . . . 60
3.5 ISAR standard field configuration . . . 64
3.6 List of instruments and their parameters on the PoB . . . 65
3.7 List of ISAR deployments from 2004 to 2012 in the English Channel and the Bay of Biscay. . . . 75
3.8 Statistics for $match - up$ pairs (AATSR CV 5 filtered dual-view SST - ISAR) from 7th December 2005 to April 2009, showing the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the range of sea temperatures spanned by the match-up database. From Wimmer *et al.* (2012) . . . 76

4.1 List of Uncertainties for ISAR . . . 81
4.2 List of parameter ranges used in the SST model . . . 84

5.1 Continuous Plankton Recorder (CPR) - Sea-Bird Electronics model 48 hull contact thermometer (SBE 48) bias correction data. The label D marks the deployment number (see table 3.7) and date is shown in day of year. A warm sky brightness temperature shows a cloudy sky and a cold sky brightness temperature a clear sky. . . . 106
5.2 Data labelling used throughout this chapter . . . 107
5.3 The mean and standard deviation of the differences for all four AATSR data products. . . . 109
5.4 Three way uncertainty analysis results for AATSR, ISAR and SBE 48 (labelled as SST_{depth}) for the years 2005 to 2008. Columns 3,4 and 5 show the sensors uncertainty. . . . 110

6.1 Match-up windows for AATSR validation . . . 116
6.2 Threshold values of indicators, I_x, for assigning quality flags Q_x . . . 131

6.3 Quality indicator results for 2004 to 2009 summary table. Table labels starting with B can be found in appendix B and table labels starting with a 6 can be found in this section. . . . 133
6.4 Quality indicator results for 2004 to 2009 for AATSR CV 5 data and robust statistics. . . . 135
6.5 Comparison between QI 3 results and Wimmer *et al.* (2012) grade 2b results for CV 5 and robust statistics. . . . 147
6.6 Comparison between QI 4 results and Wimmer *et al.* (2012) grade 1 results for CV 5 and robust statistics. . . . 147

B.1 Quality indicator results for 2004 to 2009 summary table. Table labels starting with B can be found in appendix B and table labels starting with a 6 can be found in section 6.6. . . . 163
B.2 Quality indicator results for 2004 to 2009 for AATSR data and normal statistics. . . . 165
B.3 Quality indicator results for 2004 to 2009 for AATSR CV 3 data and normal statistics. . . . 166
B.4 Quality indicator results for 2004 to 2009 for AATSR CV 4 data and normal statistics. . . . 167
B.5 Quality indicator results for 2004 to 2009 for AATSR CV 5 data and normal statistics. . . . 168
B.6 Quality indicator results for 2004 to 2009 for AATSR all data and robust statistics. . . . 169
B.7 Quality indicator results for 2004 to 2009 for AATSR CV 3 data and robust statistics. . . . 170
B.8 Quality indicator results for 2004 to 2009 for AATSR CV 4 data and robust statistics. . . . 171

C.1 Confidence values for GHRSST level 2 format with added ancillary data (L2p) *SST* values and their intended meaning . . . 215
C.2 The Single sensor error statistics (SSES) limits for AATSR . . . 216
C.3 The SSES values for AATSR. Values for TU2, TL2, TU3 and TL3 can be found in table C.2. D - N is the difference between the dual view and the nadir view temperature measurement. . . . 216

Declaration of Authorship

I, Werenfrid Wimmer, declare that the thesis titled, 'Variability and uncertainty in measuring sea surface temperature' and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- Parts of this work have been published as:

  W. Wimmer and I. S. Robinson (2008), *Report on WP 700: A New Method for Quality Assessment of SST Validation Match-ups*, National Oceanography Centre, Southampton.

Signed:

Date:

Acknowledgements

I would like to thank the following people who made this work possible:

Ian Robinson, as my supervisor, for his support, advice, encouragement and above all confidence in my work. I would also like to thank him for finding the funding for this *Ph.D.*

Craig Donlon for his advice and encouragement in the early days of this *Ph.D.*, but especially for letting me use and modify the ISAR instrument, which is his design.

Peter Challenor as the panel chair, while always encouraging he also kept the project on track scientifically and of course any errors in the statistical methods are mine.

Gary Fisher for his help and countless hours on fixing instruments, deployments, building ISARs and all sorts of technical problem solving. My gratitude also extends to the rest technical staff of the Ocean and Earth sciences department, in particular the late Ray Collins, without whom the mechanical challenges thrown up by the ISAR would have been so much harder to solve, Kevin Padley who supported the instrument deployment, Neil Jenkinson for letting us use his workshop, John Hewitt and last but by no means least Raymond Holmes the newest addition to the ISAR team.

Dave Poulter for teaching me how to process AATSR data, the shared enthusiasm of using python and of course for all those hours spent at the Pub.

Gary Corlett for discussions about the Space agency requirements for *SST* validation data in his role as AATSR validation scientist. Tim Nightingale for discussions about infra-red radiometry and instrument design. The GHRSST science team and in particular the ST-VAL team for encouragement and comments on the validation methods and procedures.

All those who kept me sane(ish) over the last few years - the JRC coffee club with too many people to mention, thank you for all the laughter and cake. And of course all my friends.

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

T.S. Eliot

The pioneer scientist must have
a vivid intuitive imagination,
for new ideas are not generated by deduction,
but by artistically creative imagination.

Max Planck

Acronyms

AATSR	Advanced Along Track Scanning Radiometer
ADC	Analogue Digital Converter
AERI	Atmospheric Emitted Radiance Interferometer
AMSR-E	Advanced Microwave Scanning Radiometer - EOS
ARC	(A)ATSR Re-analysis for Climate
ASTER	Advanced Spaceborne Thermal Emission Reflection Radiometer
ATSR	Along Track Scanning Radiometer
BB	black body
BNL	Brookhaven National Laboratory
BT	Brightness Temperature
BT_{sky}	Brightness Temperature of the sky view
BT_{sea}	Brightness Temperature of the sea view
CASOTS	Combined Action for the Study of the Ocean Thermal Skin
CASOTS II	Combined Action for the Study of the Ocean Thermal Skin second generation black body
CDR	climate data record
CEOS	Committee on Earth Observation Satellites
CH	Challenor (2004) uncertainty estimation
CIRIMS	Calibrated InfraRed In situ Measurement System
CMG	course made good
CPR	Continuous Plankton Recorder
CpF	Cap Finistere
CTD	Conductivity, Temperature, and Depth probe

CV	confidence value
D2	AATSR dual view two channel (day) product
D3	AATSR dual view three channel (night) product
DC	direct current
DECC	Department of Energy and Climate Change
Defra	Department for Environment, Food and Rural Affairs
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EM	expectation maximization algorithm
Envisat	ESA Environmental Satellite
ESA	European Space Agency
GCOS	Global Climate Observing System
GHRSST	Group for High Resolution Sea Surface Temperature
GOOS	Global Ocean Observing System
GPS	Global Positioning System
HADISST	Hadley Centre Sea Ice and Sea Surface Temperature data set
ICSU	International Council for Science
INTRA	A coupled one-dimensional air-sea interface transfer model
IOC	Intergovernmental Oceanographic Commission
IPCC	Intergovernmental Panel on Climate Change
ISAR	Infrared Sea surface temperature Autonomous Radiometer
JAXA	Japan Aerospace Exploitation Agency
KT15	Heitronics KT15.85D
L2p	GHRSST level 2 format with added ancillary data
MAD	median absolute deviation
M-AERI	Marine Atmospheric Emitted Radiance Interferometer
MLE	Maximum likelihood estimator
MS	Measurement Specialties
MSE	mean square error
N2	AATSR nadir view two channel (day) product

N3	AATSR nadir view three channel (night) product
NIST	National Institute of Standards and Technology, USA
NOCS	National Oceanography Centre, Southampton
NPL	National Physical Laboratory, Teddigton, UK
NWP	Numerical Weather Prediction
OC	O'Carroll *et al.* (2008) uncertainty estimation
OI	Optimal Interpolation
OSTIA	Operational Sea Surface Temperature and Sea Ice Analysis
Ph.D.	philosophiae doctor
PNI	PNI roll and pitch sensor
PoB	Pride of Bilbao
P & O	Peninsular and Oriental Steam Navigation Company
QA	quality assurance
QA4EO	Quality Assurance for Earth Observation
QI	quality indicator
Q-Q plot	quantile-quantile plot
RA-2	Radar Altimeter on ESA Environmental Satellite (Envisat)
RAL	Rutherford Appelton Laboratory
RSD	robust standard deviation
RSMAS	Rosenstiel School of Marine and Atmospheric Science, University of Miami
SAHFOS	Sir Alister Hardy Foundation for Ocean Science
SBE 48	Sea-Bird Electronics model 48 hull contact thermometer
SESR	surface-emitted surface-reflected radiation
SI	International System of Units
SISTeR	Scanning Infrared Sea Surface Temperature Radiometer
SOG	speed over ground
SOO	Ships of Opportunity
SOOP	Ships of Opportunity program
SSES	Single sensor error statistics

SSH	sea surface height
SST	Sea Surface Temperature
SST_{depth}	Sea Surface Temperature at depth
$SST_{foundation}$	pre-dawn Sea Surface Temperature at depth free from any diurnal variability
SST_{int}	Theoretical temperature at the air-sea interface
SST_{skin}	Sea Surface Temperature at the Skin interface
$SST_{subskin}$	Sea Surface Temperature at the base of the conductive sub-layer
std	standard deviation
Type A	Type A uncertainties
Type B	Type B uncertainties
TCM	Tokmakian and Challenor (1999) mean square error estimation
TOA	Top of Atmosphere
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational Scientific and Cultural Organization
UNFCCC	United Nations Framework Convention on Climate Change
VAR	Magnetic variation
VOS	Voluntary Observing Ships
WMO	World Meteorological Organization
XBT	Expendable Bathythermograph
YSI	Yellow Springs Instrument Company

Symbols

Symbol	Property represented	Unit
b_i	bias	
B	Brightness temperature	K
BT_{sea}	Brightness temperature of the sea	K
BT_{sky}	Brightness temperature of the sky	K
c	Speed of light	$m\ s^{-1}$
C	Counts	V
C_{BB1}	Detector counts at BB 1 view	V
C_{BB2}	Detector counts at BB 2 view	V
C_{sea}	Detector counts at sea view	V
C_{sky}	Detector counts at sky view	V
e_i	random error	
f_w	ISAR internal BB emissivity weighting factor	
g	Detector gain factor	
G	ISAR mirror gain	
G_0	ISAR pre-deployment calibration mirror gain	
h	Planck constant	$W\ s^2$
h_s	Instrument height from the sea surface	m
I_{P1}	First spatial variability indicator	
I_{P2}	Second spatial variability indicator	
I_S	Spatial mismatch indicator	
I_T	Temporal mismatch indicator	
k	Boltzmann constant	$W\ s\ K^{-1}$
L	Black body spectral radiance	$W\ m^{-2}\ sr^{-1}\ m^{-1}$
L_{amb}	Spectral radiance of ambient temperature	$W\ m^{-2}\ sr^{-1}$
L_{BB1}	Spectral radiance of BB 1 view	$W\ m^{-2}\ sr^{-1}$
L_{BB2}	Spectral radiance of BB 2 view	$W\ m^{-2}\ sr^{-1}$
L_d	Spectral radiance at the detector	$W\ m^{-2}\ sr^{-1}$
L_{down}	Down welling spectral radiance	$W\ m^{-2}\ sr^{-1}$

L_T	Spectral radiance of the target	$W\ m^{-2}\ sr^{-1}$
L_{sea}	Spectral radiance of sea view	$W\ m^{-2}\ sr^{-1}$
L_{sky}	Spectral radiance of sky view	$W\ m^{-2}\ sr^{-1}$
p	stray radiance fraction	
Q_E	Latent heat flux	$W\ m^{-2}$
Q_H	Sensible heat flux	$W\ m^{-2}$
Q_{LW}	Long wave heat flux	$W\ m^{-2}$
Q_N	Net heat flux	$W\ m^{-2}$
Q_{SW}	Short wave heat flux	$W\ m^{-2}$
Q_{P1}	QI for I_{P1}	
Q_{P2}	QI for I_{P2}	
Q_S	QI for I_S	
Q_{sky}	QI for the cloud cover of the $match - up$.	
Q_T	QI for I_T	
S_{sea}	Sea's radiance response	$W\ m^{-2}$
S_{sky}	Sky's radiance response	$W\ m^{-2}$
T_A	Air temperature	K
T_{BBamb}	Temperature of the ambient BB	K
T_B	Bulk temperature	K
T_S	Oceanic skin temperature	K
u_I	Instrument uncertainty	
u_M	Measurement uncertainty	
u'_M	Reduced measurement uncertainty	
u_P	Point in sampling uncertainty	
u_S	Spatial mismatch uncertainty	
u_T	Temporal mismatch uncertainty	
u_Z	Sampling depth uncertainty	
U	Wind speed	$m\ s^{-1}$
V_{ij}	Variance between i and j	
x_i	Observation at point i	
x_I	Observation by the *in situ* instrument	
x_T	Observation's true value	
x_S	Observation by the satellite	
X_{sea}	Fraction sea counts	
X_{sky}	Fraction sky counts	
ΔT	Temperature difference	K
ε_B	Emissivity of the sea water	
ε	Emissivity	
ε_{BB}	Emissivity of the ISAR internal BB	
ε'	Perceived emissivity change of the ISAR internal BB due to mirror contamination	

ε_0	Base emissivity for the Niclos *et al.* (2009) model	
λ	wave length	m
σ	standard deviation	
σ_B	Stefan-Boltzmann constant	$W\ m^{-2}\ K^{-4}$
Θ	viewing angle	$degrees$
τ_a	surface stress in air	
τ_{path}	path length	m
τ_w	surface stress in water	
ζ	Detector filter response	

Chapter 1

Introduction

1.1 Preface

Sea Surface Temperature (*SST*) is an important parameter in any climate monitoring system. The Global Climate Observing System (GCOS) has recognised the importance of *SST* in the Earth's climate system and designated *SST* one of the essential climate variables (ECVs) for the ocean domain. GCOS is a joint undertaking of the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational Scientific and Cultural Organization (UNESCO), the United Nations Environment Programme (UNEP) and the International Council for Science (ICSU). The aim of GCOS is to provide comprehensive information on the total climate system.

With the advent of satellites and the measurement of oceanographic parameters using satellite sensors in the 1980's it became possible to produce almost daily global *SST* maps. At the beginning these *SST* measurements were not as accurate as *in situ* measurements, although they were able to reveal the synoptic spatial variability of *SST* which is not easy to observe using instruments at sea level. Technological improvements and a better scientific understanding enabled the Along Track Scanning Radiometer (ATSR) sensor series to produce a high quality *SST* record, although that

raises the question of how the quality of the *SST* measurements is quantified. Indeed if such data are to make a contribution to the monitoring of climate change, the internationally agreed requirements for climate data records (CDRs) call for a better understanding and quantification of the uncertainties of *SST* measurements which in turn implies the need for rigour in the methods used for validation of satellite data.

The context of this thesis is a programme of validating *SST* from the Advanced Along Track Scanning Radiometer (AATSR) using the Infrared Sea surface temperature Autonomous Radiometer (ISAR) for which the author has been responsible. The new work to be described in the thesis has been associated with that programme, in particular how to exploit the unique capabilities of shipborne radiometry to improve the quality of Sea Surface Temperature at the Skin interface (SST_{skin}) measurements and to improve the precision of $match - up$ data between satellite and shipborne radiometer data. The theme of the work to be described in the thesis is to investigate the uncertainties associated with *SST* measurements, in particular instruments measuring radiometric SST_{skin}. Not only will the uncertainty budget of *in situ* SST_{skin} measurement as measured by the ISAR instrument be examined, but also the uncertainty of the $match - up$ process which ultimately contributes to the *SST* uncertainty. Furthermore, if satellite *SST* is to become a CDR not only the *SST* measurement has to be validated but also its uncertainties, for which the work performed in this study is essential.

1.2 Thesis structure

This study has been organised in the following manner:

Chapter 1 outlines the objectives of the research and shows the wider relevance of the study subject.

Chapter 2 contains a review of the research previously conducted on *SST* measurement and validation as well as providing a background to the study area.

Chapter 3 introduces the instruments and methods which produce the data used in this project. This includes a detailed description of *in situ* radiometric measurements as made by ISAR. Also a short description is given of the AATSR data together with the validation data which forms the basis for chapter 6.

Chapter 4 develops an uncertainty model for the retrieval of *SST* from a shipborne radiometer, as applied in particular to the ISAR instrument, which estimates the uncertainty value for every SST_{skin} data point.

Chapter 5 investigates how the individual uncertainties of three independent *SST* measurements, physically different but nearly coincident in space and time, can be resolved from comparisons between them using three-way uncertainty analysis. The three-way analysis presents a systematic way to verify whether SST_{skin} or Sea Surface Temperature at depth (SST_{depth}) are the better data set for validating satellite data.

Chapter 6 analyses the additional validation uncertainty introduced by the shortcomings in the matching of samples from different instruments as part of the validation process. It presents a new method for estimating the validation process uncertainty and develops the concept of a match-up quality indicator that may be applied more generally to satellite validation using *in situ* comparisons.

Chapter 7 summaries the key achievements from the previous chapters and presents the main conclusion of the study. Also possible avenues to integrate and improve the research with the wider subject area are discussed.

After the main section of the study there are a number of appendices as outlined below. The reason for the fairly large number of appendices is that the study produced quite a lot of data, which has been included for completeness but it was felt that it would disrupt the flow of the main text too much if not relegated to an appendix..

Appendix A holds the extra plots for chapter 5.

Appendix B is an extension of the results section in chapter 6.

Appendix C is a short summary of the AATSR Single sensor error statistics (SSES) model as developed by Corlett (2007).

Appendix D shows a copy of the Wimmer *et al.* (2012) paper. While this paper did not directly come from the research presented in this study it is one of the starting points for the work presented here.

1.3 The relevance and importance of *SST*

SST is one of the physical properties describing the ocean's upper structure and the air-sea interface. It therefore is not only an important parameter in furthering the scientific understanding of the oceans but also a key variable in climate studies which is reflected by GCOS specifying *SST* as an ECV (GCOS, 2010). The importance of a knowledge of *SST* for understanding our ocean is also demonstrated by *SST* being a key parameter in ocean, atmospheric, climate and Numerical Weather Prediction (NWP) models (e.g. Davies *et al.*, 2005; Kirtmann and Vecchi, 2011; Sexton *et al.*, 2003; Smith *et al.*, 2005; Storkey *et al.*, 2010). The importance of having a well characterised *SST* field was demonstrated by, amongst others, Chelton (2005) showing what effect different *SST* products have on the European Centre for Medium-Range Weather Forecasts (ECMWF) wind stress fields. Not only did Chelton (2005) find a change in magnitude of the wind stress field but also a change in the time lag, as a consequence of changing the source of *SST* data used in the NWP models. Donlon *et al.* (2012) showed the benefits of a *SST* Optimal Interpolation (OI) analysis (Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA)) on a forecasting model as demonstrated by the example of the 2007 polar ice low which was missed by most NWP models.

To be able to judge which *SST* is the best for the desired application not only the *SST* fields are required but also an assessment of its accuracy, the uncertainty, must be quantified. There have been various studies to assess the errors and uncertainties of *in situ SST* measurements such as Kent and Challenor (2006), Kennedy *et al.* (2011) and Rayner *et al.* (2006) with Kennedy *et al.* (2012) using AATSR data to asses the quality of the *in situ* data. It is noteworthy that Kennedy *et al.* (2012) shows a reverse

of previous approaches where *in situ* *SST* data were used to characterise the uncertainty of the satellite derived *SST*. For this reason the Group for High Resolution Sea Surface Temperature (GHRSST) has specified in its GHRSST level 2 format with added ancillary data (L2p) *SST* data product the need for a pixel per pixel uncertainty, referred to as the SSES.

However for most satellite *SST* sensors the SSES is estimated against a background *SST* field derived by using drifting buoys. The examples cited above show that there is potential risk of using data in the uncertainty analysis which has already been used to calibrate or otherwise influence the primary *SST* data. Minnett and Corlett (2012) consider an approach which avoids the risk by using *in situ* radiometers to assess not only the quality of the *SST* but also the uncertainty model used for the satellite sensor. The use of measurements which are traceable to a common standard, such as satellite sensors and *in situ* radiometers makes it possible to adopt the Minnett and Corlett (2012) approach for validating satellite derived *SST* CDR. The Minnett and Corlett (2012) approach also satisfies the Quality Assurance for Earth Observation (QA4EO) requirement for a) comparing only those measurement which are traceable to the same standard and b) a quality indicator (based on uncertainty analysis) assigned to each measurement (Fox, 2010).

1.4 Aims and Objectives

The aim of this study is to investigate the factors that contribute to the accuracy of *SST* measured from ships and the effect that such *in situ* measurements can have on their use for validating satellite-derived *SST* data products, particularly in relation to using the satellite data as CDR. This is especially relevant in integrated measurement systems, where measurement uncertainties must be well controlled since each component could have a detrimental effect on the whole system. Now that there are over thirty years of satellite *SST* data available from which potentially a CDR quality satellite *SST* record can be produced, not only do the *SST* measurements require validation but there is also

a need to determine how reliable are the estimates of uncertainty. (Minnett and Corlett, 2012).

The specific objective of this study is to **characterize the uncertainty associated with measurements of SST_{skin} and SST_{depth}**. In particular the focus will be on three main areas:

A.1 Assessing the uncertainties associated with SST_{skin} measurements systems, using ISAR as an example.

A.2 Evaluating the uncertainties related to different measurement techniques on board Voluntary Observing Ships (VOS) and Ships of Opportunity (SOO), especially in relation to SST_{depth} and SST_{skin} measurements. This includes the use of advanced uncertainty calculation methods developed by, for example, Tokmakian and Challenor (1999) and O'Carroll *et al.* (2008).

A.3 Examining the uncertainties introduced by validation methods, which permit some latitude in the specification of the spatial and temporal coincidence of samples from different datasets being inter-compared. The focus here is on the uncertainty budget of the $match - up$ process.

The scientific questions that lie behind the three objectives are:

Q.1 How do conventionally measured SST and its uncertainties impact the climate record? For this study the "conventionally measured" SST refers to shipborne SST_{depth} measurements.

Q.2 Can SST_{skin} measurements acquired by satellites be a better data source for a climate record than buoy measured SST_{depth} ?

Q.3 To what extent has a poor choice of validation data and validation method led to under-appreciation of the quality of satellite acquired SST_{skin}.

The main objectives have been conceived in order to answer the scientific questions and to provide a structured approach to the research carried out in this thesis. Chapter 7 will point to how the scientific questions raised here can be addressed using the work presented in the earlier chapters.

Chapter 2

Background

The research work presented in this thesis builds on the existing knowledge of *SST* measurements in attempting to advance the understanding of the uncertainties associated with *SST* measurements. Therefore it is appropriate in this chapter to summarise and assess the current state-of-the-art concerning *SST* measurements and to consider why a better understanding of *SST* is important in a climatological and operational context.

First the upper ocean temperature structure and why this structure is important for our understanding of satellite *SST* measurements is discussed. Next the thermal skin effect at the sub-centimetre scale is explained and how it depends on air-sea fluxes. This is followed by a section on the theoretical principles governing the radiometric measurements of SST_{skin}, followed by a section on diurnal variability and why it is important for the accurate use of *SST* measurements. Then the importance of *SST* as an ECV is shown, followed by a discussion of the operational requirements to meet measurement protocols for climate. Finally a short overview of *in situ* radiometers is provided, together with an overview of the geographical study area within which the data were measured.

2.1 Temperature in the upper ocean

When physical oceanographers refer to the upper ocean temperature they generally mean the water temperature above the thermocline, the temperature of the mixed layer, so called because wind stirring and wave action are assumed to break down vertical stratification near the the surface producing vertical uniformity of temperature and salinity throughout the layer. The mixed layer depth can vary from a few tens of metres to a few hundred metres depending on the season and the location. The main factors controlling the varying mixed layer depth are the heat flux and the wind speed. On sunny days with very little wind the very top of the ocean develops a so called diurnal thermocline where the ocean is heated by the solar radiation and the temperature in the top few meters is raised above the mixed layer temperature (reviewed in more detail in section 2.4). Figure 2.1 shows two characteristic extreme cases of upper ocean temperature profiles. Labelled as a) in figure 2.1 is the profile usually found at night time and also during the day if the wind speed is high. In this case the temperature is uniform with depth from tens of metres deep in the mixed layer up to within a few mm of the surface where a cool skin layer is found (the cool skin effect is reviewed in more detail in section 2.2). Labeled as (b) is a day time high insolation and low wind speed temperature profile where a temperature "step" of up to 6 K (Stuart-Menteth *et al.*, 2003) can be found under certain conditions.

The labels used to refer to the temperatures at different points within the profile in figure 2.1 are those defined by Donlon *et al.* (2007), and are the definitions used by the Group for High Resolution Sea Surface Temperature (GHRSST). The SST_{int} corresponds to the temperature of water and air molecules precisely at the liquid-gas interface. It must be regarded as a theoretical variable because it can not be measured with current (2012) technology. The term SST_{skin} is used to represent the temperature within the top few microns of the sea, where flux of sensible heat towards the boundary with the atmosphere creates a cool skin (Saunders, 1967). Formally the variable SST_{skin} is defined as the temperature measured by infra-red radiometers operating in the waveband 3.7 - 12 μm , which represents the temperature within the conductive diffusion dominated sub layer

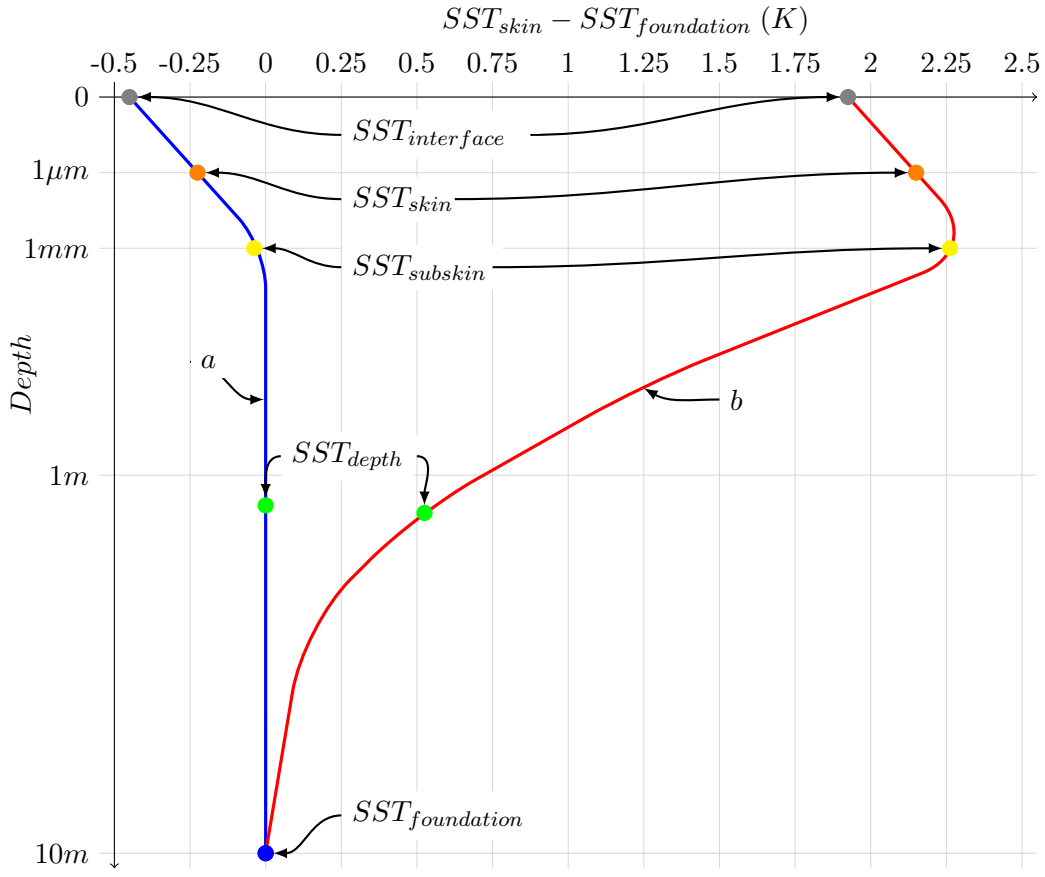


FIGURE 2.1: Schematic view of the upper ocean temperature profile representing a) night-time and day time high wind conditions in blue and b) day time low wind conditions in red. Note that the temperature differences as given are typical but in reality can vary from these values quite strongly. Values of (SST_{skin} - $SST_{foundation}$) of up to 6 K have been recorded. Also note that the SST_{depth} corresponds to a SST measurement made by a thermometer in the water column, which could be at any depth from 0 depth to the bottom of the sea, although for this study SST_{depth} refers to as a temperature somewhere between $SST_{subskin}$ and $SST_{foundation}$. After Donlon *et al.* (2007)

at depths of 10 - 20 μm (Donlon *et al.*, 2007). The Sea Surface Temperature at the base of the conductive sub-layer ($SST_{subskin}$) corresponds to what the SST would be in the absence of the cool skin. For practical purposes $SST_{subskin}$ can be well approximated to the measurement of SST by a microwave radiometer operating in the 6 - 11 GHz frequency range. However this approximate equivalence is neither direct nor invariant to changing physical conditions or to the specific geometry of the microwave measurements.

Measurements at depths below $SST_{subskin}$ are referred to as SST_{depth}. SST_{depth} measurements can be made by a variety of sensors mounted on ships, buoys and fixed platforms and range in sampling depth from a few mm to the core of the upper mixed

layer tens of metres below the surface. These measurements are very different from the temperatures measured by remote sensing techniques (SST_{skin} and $SST_{subskin}$) and therefore must be qualified by assigning a depth in metres to each measurements (e.g.: SST(5m)). Finally the label of $SST_{foundation}$ is used to represent the temperature free from any diurnal variability. Minnett (2011) defines $SST_{foundation}$ as the temperature at the first time of the day when the heat gain from the solar radiation absorption exceeds the heat loss at the sea surface. Stuart-Menteth (2004) showed that $SST_{foundation}$ is the temperature value to which temperature at all near surface depths collapses before sunrise. From a practical standpoint the direct measurement of the $SST_{foundation}$ is not easily achievable with satellite measurements. However from Stuart-Menteth (2004) the $SST_{foundation}$ can be approximated by the measurement of $SST_{subskin}$ at dawn before any diurnal heating can have occurred. Using satellite measurements $SST_{foundation}$ is therefore directly measurable just once per day. Note that, as a proxy for the mixed layer temperature (assumed uniform with depth) there is no question of assigning a depth to it.

During a diurnal warming event there is no way that remote sensing can determine the depth of the diurnal thermocline, but it is assumed that by the following dawn the diurnal thermocline has been destroyed by convective overturning and any net change in heat content has been redistributed through the mixed layer.

Historically what is now defined as $SST_{foundation}$ has been referred to as the bulk temperature. However the term bulk temperature has also been used for any temperature measurement in the mixed layer. Without a depth associated to it, and without any knowledge of the coincident wind or insolation conditions, such a "bulk" SST measurement contains considerable uncertainties for interpreting how it can be used for validating satellite SST observations or in assembling SST climatologies (Schluessel *et al.*, 1990). The GHRSST recommendation is that in future the term bulk temperature should be deprecated. Table 2.1 gives a short summary of the SST definitions as described here (after Donlon *et al.*, 2007; Kawai and Wada, 2007).

TABLE 2.1: Definitions of sea surface temperature proposed by GHRSST (Donlon *et al.*, 2007; Kawai and Wada, 2007)

Name	Symbol	Temperature represented	Instrument for measuring
Interface *SST*	SST_{int}	Theoretical temperature at the air-sea interface	None
Skin *SST*	SST_{skin}	Temperature within the conductive diffusion dominated sub-layer $\approx 10-20\mu m$ depth.	Infra-red Radiometer with a 3.7-12 μm waveband.
Sub-skin *SST*	$SST_{subskin}$	Sea Surface Temperature at the base of the conductive sub-layer at $\approx 1mm$	Microwave radiometer at 6-11 Ghz, high performance autonomous profiler (Ward *et al.*, 2004)
Temperature at depth	SST_{depth}	Temperature measured at depth usually referred to as "bulk" *SST*	Most *in situ* sensors, e.g. CTD, XBT, engine intake
Foundation *SST*	$SST_{foundation}$	Temperature of water column free of any diurnal signal	same as SST_{depth}

2.2 The skin layer and the cool skin effect

The skin layer temperature (SST_{skin}) is normally a few tenth of a Kelvin cooler then the sub-skin temperature $SST_{subskin}$ (Saunders, 1967). The skin layer is the boundary layer, between a turbulent ocean and a turbulent atmosphere, in which molecular diffusion processes and molecular viscosity influence the flux rates across the layer. This molecular layer is necessary for the transfer of heat, momentum, mass and gas between the ocean and the atmosphere (Emery *et al.*, 2001).

The main driver for the skin layer temperature gradient is the net upward heat flux Q_N. Equation 2.1 shows the components for the net heat flux Q_N:

$$Q_N = -Q_{SW} + Q_{LW} + Q_H + Q_E \tag{2.1}$$

where Q_{SW} is the insolation or the flux of solar energy into the sea, Q_{LW} the infra-red radiation from the sea, Q_H the sensible heat flux and Q_E the latent heat flux. A positive sign in equation 2.1 represents a flux from the ocean to the atmosphere. The main constraints for the skin effect is the sign of $Q_{LW} + Q_H + Q_E$, since Q_{SW} is passing through the skin layer without affecting the heat budget in the top mm. Because the sign of Q_{LW} and Q_E is almost always positive, even a negative Q_H, meaning the air temperature is warmer than the ocean, will not change the sign of the net heat flux in the top mm, which is almost always positive. As a consequence of a positive net heat flux in the top mm the skin layer temperature gradient is negative, resulting in a cool skin. The cool skin temperature difference (SST_{skin} - $SST_{subskin}$, ΔT) is on average -0.17 ± 0.07 K (Donlon *et al.*, 2002). Because measurements of the skin layer were very sparse before infra-red radiometers existed (Hasse, 1963; Saunders, 1967) various models were derived to represent the cool skin effect and, where necessary, used to convert observations of SST_{skin} to estimates of $SST_{subskin}$, or observations of $SST_{subskin}$ to estimates of SST_{skin} (e.g.: Fairall *et al.*, 1996a; Gentemann *et al.*, 2009). The physical processes that govern the magnitude ΔT can vary with the environmental conditions (Emery *et al.*, 2001). Three different possible regimes are summarized schematically in figure 2.2.

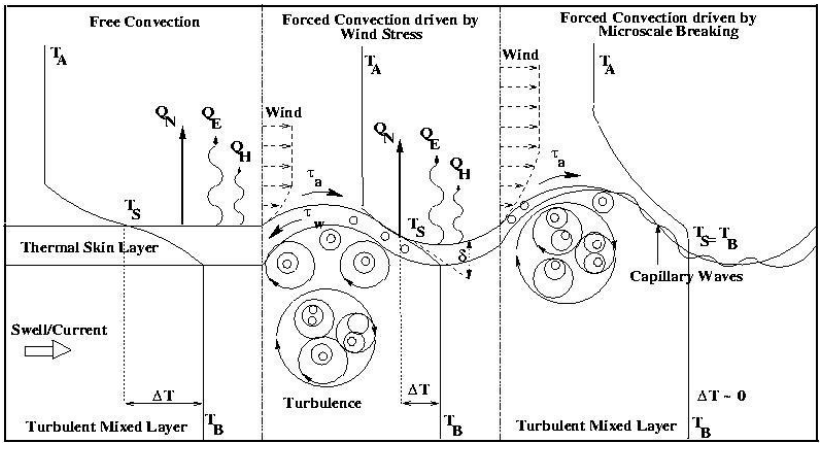


FIGURE 2.2: Behaviour of the oceanic skin layer under three primary physical regimes: T_A, T_S, and T_B represent, respectively, the air temperature, oceanic skin temperature, and bulk ocean temperature; Q_E, Q_H, and Q_N represent the latent, sensible, and net radiative heat fluxes; and τ_a and τ_w represent the surface stress in the air and the water (from Emery *et al.*, 2001).

These regimes include free convection, forced convection driven by wind stress and forced convection driven by micro-scale wave breaking. The figure shows the primary processes

affecting ΔT in each of the three regimes and the representative conditions in the upper ocean and lower atmosphere.

Castro *et al.* (2003) state that during **free convection** the turbulent transport of heat is buoyancy driven. Since the cool skin is denser than the underlying water, it will become gravitationally unstable and tend to sink, provided that the duration of the locally unstable conditions is long compared with the characteristic time scale of the instability process. Castro *et al.* (2003) furthermore states that evaporation generates a salinity gradient that also contributes to the gravitational instability. During free convection, mean flow and wind shear stress are, by convention, assumed to be absent (they are non existent or very weak) (Castro *et al.*, 2003). As a result, ΔT is controlled primarily by the net heat flux at the surface. The free convection mechanism is always present, but different processes at the sea surface may work simultaneously, leading to more frequent renewal through increased turbulent mixing.

Forced convection (Castro *et al.*, 2003) occurs when the micro scale turbulence is sustained by mechanical forces acting on the sea surface. The two main forcing mechanisms arise from either shear flow instabilities generated by direct variations in surface wind stress, or from the interactions between the surface stress and the orbital velocities of small-amplitude surface waves directly generated by the wind such as capillary waves and short gravity waves. During the latter situation, the surface renewal is the result of density fluctuations induced by waves in which the vertical density gradient associated with the wave becomes positive, leading to convective overturning. During forced convection driven by wind shear stress, a shear instability is generated as a result of the velocity difference between the top and bottom of the viscous sub layer, leading to sporadic over turnings of the shear flow. ΔT is regulated both by the net heat flux and the wind shear stress. If the wind velocity increases as the sea surface is being cooled, there is a more rapid cooling and also a thinning of the viscous sub layer.

Figure 2.3 shows a modelled vertical temperature profile for the cool skin effect. A coupled one-dimensional air-sea interface transfer model (INTRA) model was developed by Eifler and Donlon (2001) to investigate atmosphere-ocean heat transfer processes characterized by periodic wave breaking. The INTRA model was verified with Calibrated

InfraRed In situ Measurement System (CIRIMS) data (Jessup *et al.*, 1997) for a number of conditions. The air temperature profiles in figure 2.3 and figure 2.2 are quite different, and while both profiles are possible, the profile in figure 2.3 shows the more common air temperature profile.

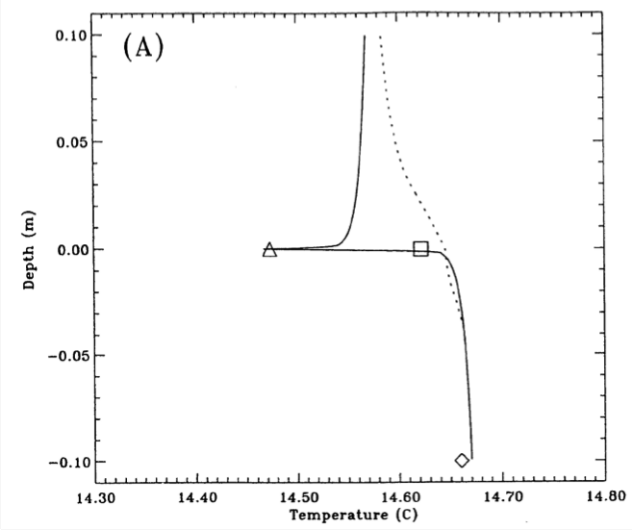


FIGURE 2.3: Vertical temperature profiles between 0.1 m above(atmosphere) and 0.1 m below (ocean) computed by the INTRA model Eifler and Donlon (2001). Triangles represent observed mean SST_{skin}; squares represent maximum SST_{skin} and diamonds represent SST_{depth}.

The cool skin effect is persistent and has to be considered in any SST_{skin} to SST_{depth} comparison. This is especially true for satellite SST_{skin} validation where the uncertainty introduced by not accounting for the cool skin effect can represent a considerable part of the uncertainty budget, for example for AATSR the total uncertainty is 0.3 K. For satellite validation the cool skin effect can be cancelled out by using *in situ* radiometers which are measuring SST_{skin}.

2.3 Radiometric temperature measurements

The basis of all radiometric temperature measurements is that all surfaces emit infra-red radiation. The spectral characteristics of the thermal emission for a black body (BB) are specified by the Planck equation 2.2.

$$L(T,\lambda) = \frac{2hc^2}{\lambda^5 \left[e^{(\frac{hc}{\lambda kT})} - 1\right]} \tag{2.2}$$

where λ is the wavelength in metres, L the spectral radiance and k is the Boltzmann constant with $k = 1.3806488 \cdot 10^{-23}\ W\ s\ K^{-1}$, h is the Planck constant with $h = 6.62606957 \cdot 10^{34}\ W\ s^2$, c is the speed of light with $c = 2.99792 \cdot 10^{8}\ m\ s^{-1}$ and T the temperature of the thermal emission in Kelvin. Integration over all wave lengths gives the ideal black body radiation:

$$L = \sigma_B T^4 \tag{2.3}$$

where σ_B is the Stefan-Boltzmann constant with $\sigma_B = 5.670373 \cdot 10^{-8}\ W\ m^{-2}\ K^{-4}$. Figure 2.4 shows the Planck function in the infra-red part of the spectrum for a range of temperatures. While equations 2.2 and 2.3 represent the black body radiation, the sea surface is not a perfect emitter. The emissivity accounts for the sea surface not being a perfect emitter and is defined as the ratio of the emission of a non perfect emitter to the emission emitted by a black body at the same temperature. Equation 2.4 shows how the emissivity is defined, with $\varepsilon(\lambda)$ representing the emissivity and $L(\lambda)$ the emission of a BB.

$$\varepsilon(\lambda) = \frac{Emission\ of\ a\ real\ surface\ at\ \lambda}{L(\lambda)} \tag{2.4}$$

The emission peak of the infra-red radiation lies typically around 9 μm to 11 μm, which makes this part of the thermal infra-red spectrum ideal for SST_{skin} measurements from space, especially as there is a spectral atmospheric window located at 10.0 to 12.5 μm. An infra-red radiometer detects the Brightness Temperature (BT) of the infra-red emitting surface which is the equivalent temperature of a black body emitting across the detector's waveband.

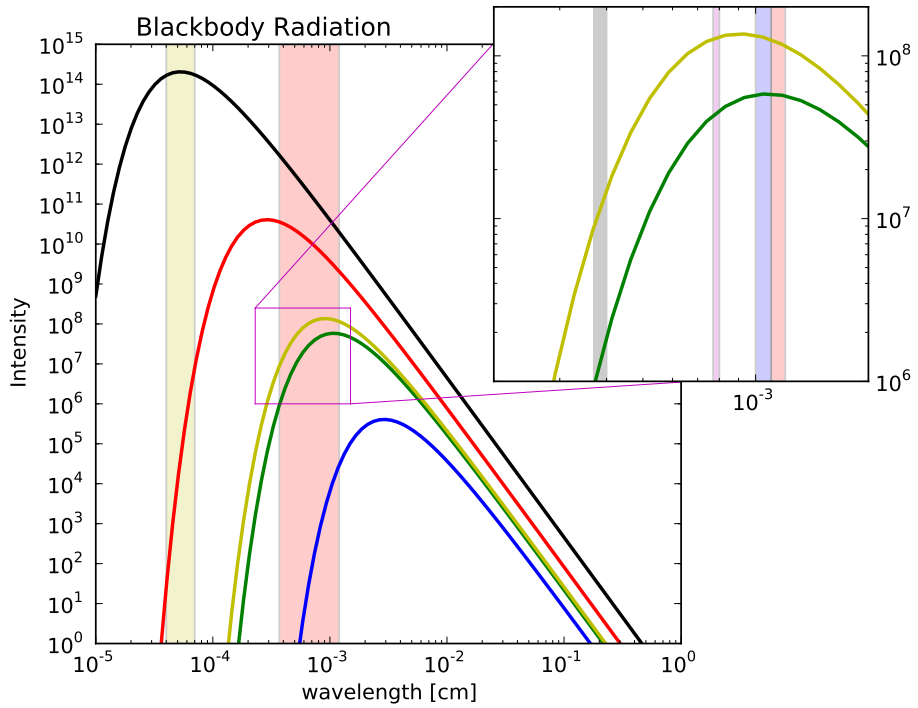


FIGURE 2.4: Black body radiation. The blue line represents the thermal emission for a temperature of 100 K, the green line 270 K, the yellow line 320 K, the red line 1000 K and the black line is approximately the temperature of the sun of 5500 K. The two shaded areas show in yellow the visible spectrum and in red the infra-red spectrum between 3.7 and 12 μm in the main plot. The zoomed out section in the top right corner of the plot show again 270 K and 320 K as green and yellow lines. The shaded areas are four window regions used for *SST* measurements from satellites. Shaded grey is 3.7 μm , magenta is 7.7 μm , blue 11 μm and red 12 μm .

Sea Surface Emissivity (ε) has a value of slightly less than one, but it is not constant and has not only a dependence on viewing angle and the sea state, but also on *SST* and surface salinity (Newman *et al.*, 2005). A number of theoretical models exist for estimating ε, with the most commonly used ones being those from Masuda *et al.* (1988) and Wu and Smith (1997). The form of the latter is shown in figure 2.5. The (A)ATSR Re-analysis for Climate (ARC) project, for example uses an emissivity model based on the work from Watts *et al.* (1996), Masuda *et al.* (1988) and Wu and Smith (1997), with some improvements from Newman *et al.* (2005). The improvements introduced by Newman *et al.* (2005) incorporate a better model for the temperature dependence of ε based on interferometer measurements. The Masuda *et al.* (1988) and Wu and Smith (1997) models show that with increasing wind speed ε decreases. However Hanafin and Minnett (2005) and Niclos *et al.* (2005) found from observations that this only applies

for view angles below 40°. At higher view angles ε increases with increasing wind speed. The higher dependence of ε on the sea state over a 40° viewing angle coincides with a sharp drop in ε. As a result Masuda (2006) revisited the Masuda *et al.* (1988) and Wu and Smith (1997) calculations and added a surface-emitted surface-reflected radiation (SESR) term which results in ε being similar to Hanafin and Minnett (2005). The SESR term has very little effect below a 40° view angle and increases from a view angle of 50° showing a maximum increment to ε of 0.03 at a view angle of 80°.

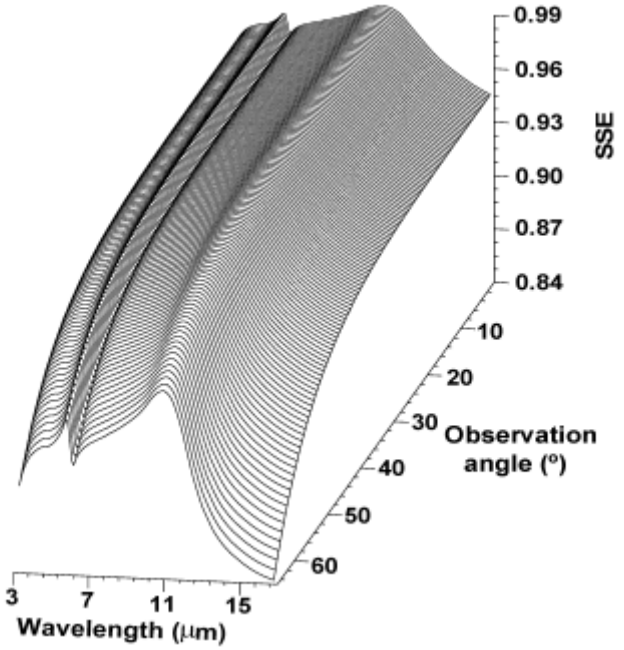


FIGURE 2.5: ε (labelled as SSE) dependence on wavelength and observation angle, as given by the model of Wu and Smith (1997) for a wind speed $U = 0ms^{-1}$, from Niclos *et al.* (2009)

For *in situ* radiometers ε has a big influence on the calculation of SST_{skin}. This is especially true because, in most cases, ε is not estimated from the *in situ* radiometers measurements itself but used as a predetermined constant, which might not be correct in all measurement conditions. Section 4.2.3.1 evaluates the uncertainties a suboptimal ε introduces into the SST_{skin} calculation.

2.4 Diurnal variability

During the day the upper ocean temperature is strongly dependant on the insolation and wind speed. As shown in figure 2.1 the temperature profile, while the SST_{skin} effect persists, can have a surface temperature that is several K warmer than a few m deeper because of solar heating in low wind conditions. The diurnal variability was first recognised by Sverdrup *et al.* (1942) and subsequently Stommel (1969) and has since been studied extensively to quantify the the effect (Donlon *et al.*, 2002; Fairall *et al.*, 1996b, 2003; Gentemann and Minnett, 2008; Gentemann *et al.*, 2009; Stuart-Menteth *et al.*, 2003, to name a few). Price *et al.* (1986) showed that under summer low wind conditions the solar heating extends to approximatively 1 m beneath the surface. Under light and moderate wind conditions the heated depth can extend to around 10 m below the surface due to increased vertical mixing and can persist for a few hours (Minnett, 2003).

The main reasons for studying the diurnal variability is that, if ignored, significant errors are introduced into the surface fluxes of heat and momentum in NWP and climate models (Webster *et al.*, 1996). Donlon and Robinson (1997) for example report that the error in the total heat budget can exceed $100 W m^{-2}$ when SST_{depth} measurements are used for flux comparisons rather than SST_{skin}.

In the context of validation diurnal variability is together with cloud mask issues one of the biggest sources of uncertainties. Comparing satellite SST_{skin} with *in situ* SST_{skin} can reduce the effect diurnal variability has on the validation result. However it does not cancel out the effect diurnal variability has completely. Therefore great care has to be taken when designing the $match - up$ temporal window to minimise the impact diurnal variability has on validation statistics.

2.5 Climate relevance of temperature measurements

The GCOS defines 44 ECVs, of which 15 ECVs are related to the ocean. A summary of every ECV as defined by GCOS is shown in table 2.2. The ECVs are required to support

TABLE 2.2: List of the ECV as defined by GCOS

Domain		ECV
Atmospheric	Surface	Air temperature Precipitation Air pressure Surface radiation budget Wind speed and direction Water vapour
	Upper-air	Earth radiation budget Upper-air temperature Wind speed and direction Water vapour Cloud properties
	Composition	Carbon dioxide Methane Ozone Other long-lived greenhouse gases Aerosol properties
Oceanic	Surface	Sea-surface temperature Sea-surface salinity Sea level Sea state Sea ice Current Ocean colour Carbon dioxide partial pressure
	Sub-surface	Temperature Salinity Current Nutrients Carbon Ocean tracers Phytoplankton
Terrestrial		River discharge Water use Ground water Lake levels Snow cover Glaciers and ice caps Permafrost and seasonally-frozen ground Albedo Land cover (including vegetation type) Fraction of absorbed photosynthetically active radiation Leaf area index Biomass Fire disturbance Soil moisture

the work of the United Nations Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC). GCOS deems that all ECV are technically and economically feasible for systematic observation. Furthermore GCOS states the importance of international exchange not only of data but also the exchange of measurement and validation protocols for both current and historical observations.

Ohring *et al.* (2005) states that for climate studies the requirement for *SST* is an absolute uncertainty of 0.1 K and a stability of 0.04 K $decade^{-1}$. The required accuracies and stabilities for the climate variable datasets were established with regard to being able to detect changes in important climate signals based on current understanding, expectations and models of long-term climate change. They also state, as an example, that if a somewhat arbitrary stability factor of 20 % is assumed the measurements would have an uncertainty range of 0.8 to 1.2, or a factor 1.5. Ohring *et al.* (2005) compared the arbitrary stability factor of 20 % to climate models with a range of 1.4 to 5.8 K, or a factor of about 4, in the change in global temperature by 2100, and concluded that measurements with a stability of better than 20 % should greatly help to reduce the uncertainty of our climate predictions.

Achieving an absolute uncertainty of 0.1 K and a stability of 0.04 K $decade^{-1}$ is not an easy task and for example Rayner *et al.* (2009) has shown that historical *SST* can on a global average have bias adjustment uncertainties of up to 0.08 K, with local values ranging from 0.05 to 0.2 K. The main reasons for the high uncertainty is the lack of adequate meta data. Kennedy *et al.* (2011) has shown that the global annual uncertainty in the Hadley Centre Sea Ice and Sea Surface Temperature data set (HADISST) decreases from 0.08 K in the 1870's to 0.02 K in the period from 1960 to 2000. More recently, especially with the greater availability of satellite data, studies like Minnett and Corlett (2012) are trying to use satellite *SST* as a CDR. Rayner *et al.* (2009) stated that in an ideal situation *in situ* and satellite measured *SST* work together as an observation system for climate, with the satellite system having the better spatial resolution and coverage and *in situ* data being used for calibration and uncertainty verification. Rayner *et al.* (2009), amongst others, also states the importance of providing an individual uncertainty for each satellite pixel. Minnett and Corlett (2012) proposes

to address this by using *in situ* buoy and ship observations to calibrate and validate the satellite *SST* and *in situ* radiometers to validate the uncertainty model for the satellite *SST*.

2.6 Operational requirements

Operational oceanographic applications of *SST*, such as operational ocean and weather forecasting services have their own particular requirements. They need to have a timely delivery of the best *SST* available to use as a boundary condition for a forecasting model. In the case of SST_{skin} the parameter not only acts as a boundary condition for the oceanographic model but also for the atmospheric models (Martin *et al.*, 2007). However for historic reasons most oceanographic models run by the meteorological service can only assimilate SST_{depth} measurement - or ideally a pre-dawn Sea Surface Temperature at depth free from any diurnal variability ($SST_{foundation}$) which means that satellite *SST* measurement have to be converted to a SST_{depth} temperature with an upper ocean model also known as diurnal variability models (e.g. Gentemann and Minnett, 2008; Stuart-Menteth *et al.*, 2003). Furthermore because the measured *SST* often have gaps, because either there was cloud present in the case of infra-red measurement or the measurement is too close to land in the case of microwave measurements or data are missing, operational agencies run an OI scheme to get the best estimate of *SST* possible with the available data. Using OI analyses does not completely solve the issue of data gaps because normally a compromise has to be made between achieving the data gap filling by generating smooth data fields while retaining enough of the spatial structure to actually resolve the oceanographic features (Donlon *et al.*, 2012; Stark *et al.*, 2007).

The GHRSST[1] (Donlon *et al.*, 2007) has addressed some of the issues related with operational satellite *SST* use and developed a uniform data format for all satellite *SST* sensors which incorporates additional information such as wind speed, solar insolation to make the interpretation and if necessary transformation from SST_{skin} to SST_{depth} as simple as possible.

[1] www.ghrsst.org

2.7 $In-situ$ Radiometers

In chapter 3 the ISAR radiometer is introduced as a primary source of the *in situ* SST_{skin} data records used in the rest of this work. To some extent, the observation made with the ISAR instrument can be made with any infra-red *in situ* radiometer. In this section a short description of each of the three main alternatives to ISAR is given. The three radiometers mentioned here are by no means the only *in situ* radiometers reported in the literature, but these are the only other self-calibrating traceable radiometers currently used for systematic SST_{skin} observations. Figure 2.6 shows the four radiometers (ISAR, the Marine Atmospheric Emitted Radiance Interferometer (M-AERI), CIRIMS and the Scanning Infrared Sea Surface Temperature Radiometer (SISTeR)) at the Miami 2001 inter-comparison experiment. Inter-comparison between the radiometers is important to ensure a) that the measurements of SST_{skin} are comparable and b) that the radiometers are traceable to International System of Units (SI) standards (Theocharous *et al.*, 2010). The traceability is normally achieved at the inter-comparison workshops by comparison with a reference radiance source, or a reference radiometer, provided by one of the national standard agencies such as National Institute of Standards and Technology, USA (NIST) or National Physical Laboratory, Teddigton, UK (NPL). More information on the latest inter-comparison workshop (2009) can be found in section 3.2.5.3.

2.7.1 M-AERI

The M-AERI is a seagoing development of the Atmospheric Emitted Radiance Interferometer (AERI), and is a robust, accurate, self-calibrating, seagoing Fourier-transform interferometric infra-red spectroradiometer that is deployed on marine platforms to measure the emission spectra from the sea surface and marine atmosphere. The M-AERI performs spectral measurements in a range from 3 μm to 18 μm and uses two black bodies (BB), one heated and one at ambient temperature to calibrate each measurement. The instrument is traceable to NIST standard and is calibrated before and after each deployment against a NIST traceable laboratory BB. Figure 2.7 shows the M-AERI

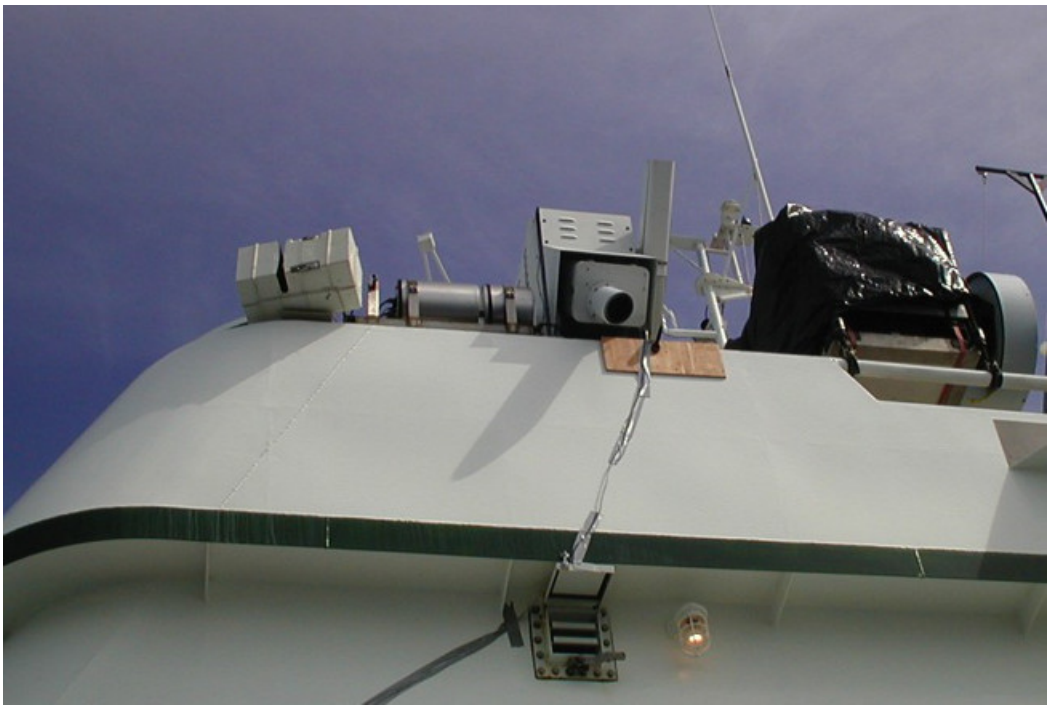


FIGURE 2.6: The Miami 2001 inter-comparison experiment. On the far right the M-AERI is visible, next is the CIRIMS, to the left of CIRIMS is ISAR 01 and far left is SISTeR. From Barton *et al.* (2004)

on the Explorer of the Seas and the instrument is described in detail in Minnett *et al.* (2001).

FIGURE 2.7: The M-AERI on the Explorer of the Seas. From Minnett (2004).

The main difference between M-AERI and the other *in situ* radiometers is that M-AERI measures the whole spectrum from 3 μm to 18 μm with a resolution of $0.5cm^{-1}$.

A number of M-AERI have been deployed over the years on research cruises all over the world as well as the Royal Caribbean Explorer of the Seas which was on a repeat track around the Caribbean from Miami to Saint Thomas from 2002 until the end of 2008. P. Minnett at the University of Miami, who is responsible for the M-AERI instruments also uses two ISARs which are deployed on various SOO.

2.7.2 SISTeR

Nightingale (2007) describes SISTeR as a compact and robust chopped self-calibrating filter radiometer. Its dimensions are approximately 20 by 20 by 40 cm and it weighs about 20 kg. The instrument is divided into three compartments containing the fore-optics, scan mirror and reference black bodies, and a small-format PC with signal processing and control electronics. The SISTeR has been designed to survive and maintain its calibration over extended periods in a maritime environment. The instrument has a DLATGS pyroelectric detector with a filter wheel containing three narrow-band filters centred at 3.7 μm , 10.8 μm and 12.0 μm , matching those in the ATSR instruments. It also has a scan mirror and a ZnSe window protecting the instrument from the elements. SISTeR has two BBs fitted, with one being at ambient temperature and one at approximately 10 K above ambient temperature. Figure 2.8 shows the SISTeR instrument. All SISTeR measurements are traceable to NPL and NIST standards.

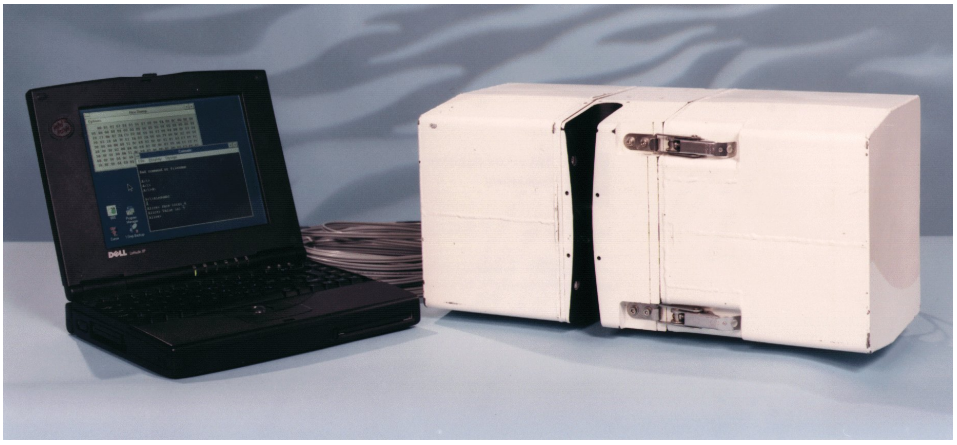


FIGURE 2.8: The SISTeR. From Nightingale (2007)

SISTeR is very similar to ISAR, with the main difference being that SISTeR was designed as a research instrument needing an operator during deployments and ISAR as a fully

autonomous instrument for deployment on SOO. SISTeR has been made autonomous in recent years and is now deployed on the Cunard liner Queen Mary 2.

2.7.3 CIRIMS

Jessup *et al.* (2002) describes CIRIMS design goals as to provide ocean skin temperature data with an accuracy of $\pm 0.1K$ from a system that has the ability to run autonomously at sea for extended periods with no involvement from the vessel crew. The CIRIMS design incorporates two Heitronics infra-red KT11 radiometers with a spectral bandwidth in the 9.6-11.5 μm range. One radiometer is housed within the unit itself and measures sea surface radiance. The second radiometer is housed externally in an enclosure and measures sky radiance. Reliable calibration of the internal radiometer is achieved by two-point calibration using a modified Hart Scientific microbath. A custom designed, copper cylindro-cone BB is immersed in a water/ethylene glycol solution within the temperature-controlled microbath. Two calibration points are set a few degrees above and below the scene temperature allowing for dynamic calibration over a wide range of scene temperatures. The temperature-controlled housing provides a stable, dry environment for the internal radiometer and the BB. The insulated housing is heated and cooled by a thermoelectric heater/cooler unit, which maintains the internal case temperature to within $\pm 0.1K$of the set point, generally $35°C$. Protection of the radiometer and BB is arguably the most challenging and debated aspect of a practical design. Jessup *et al.* (2002) chose to use an infra-red transparent window to provide complete protection under all conditions. This approach relies on the ability to correct for the effect of the window (Jessup *et al.*, 2002). This is also the main difference between CIRIMS and the other radiometers described here, while ISAR, SISTeR and M-AERI have an optical path that does not change regardless of whether the BB or the sea/sky is viewed, CIRIMS has a ZnSe window on the outside and therefore changes the optical path when the sea is viewed. Figure 2.9 shows the CIRIMS installed on the R/V Ron Brown. CIRIMS is, like the other radiometers described here, traceable to NIST.

FIGURE 2.9: The CIRIMS installed on the R/V Ron Brown. From Jessup (2003).

2.8 Measurements on Ships of Opportunity and Voluntary Observing Ships

The primary goal of the Ships of Opportunity program (SOOP) is to fulfill upper ocean data requirements which have been established by Global Ocean Observing System (GOOS) and GCOS, and which can be met at present by measurements from Ships of Opportunity (SOO) (JCOMM, 2010a). In contrast the WMO Voluntary Observing Ships (VOS) scheme relies on ships recruited by national meteorological services for acquiring and transmitting meteorological observations (JCOMM, 2010b). While normally VOS are more used for meteorological measurement and SOO for oceanographic measurements the main distinction is that VOS fulfill an operational task for the met services while SOO are somewhat more opportunistic and their routes and vessels may

be subject to frequent changes. The potential for frequent vessel changes can make measurements at fixed installations or along well defined routes difficult to maintain using SOO.

The two vessels used in this study as instrument platforms (see next chapter) were classed as VOS in the case of the Pride of Bilbao (PoB) and as SOO in the case of the Cap Finistere (CpF).

2.9 Study area

The data for this study has been collected in the Bay of Biscay and the English Channel on board the PoB (from 2004-2010) and the CpF (2010-2012) following approximately the route shown in Figure 2.10. The data collection started in March 2004 and is ongoing to date.

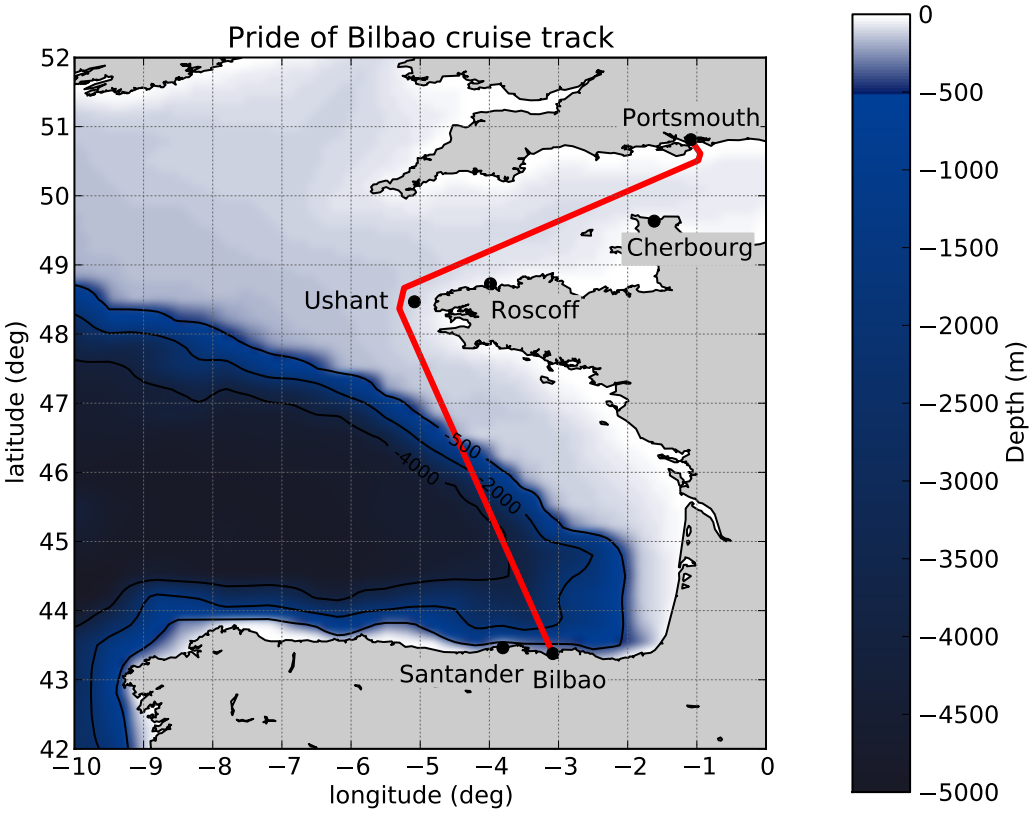


FIGURE 2.10: Overview of the study area with a typical ship track super-imposed on the bathymetry of the area

The study area can be divided in to two main water bodies, the English Channel and the Bay of Biscay. Bargeron *et al.* (2006) divided the PoB ship track in eight different regimes starting with Portsmouth Harbour, then the shallow well mixed central English Channel with depths of 30 to 60 m, continuing with the shallow well mixed, summer stratifying western Channel, followed by the western approaches near the French coast at Ushant with depths 30 to 130 m, going into the Bay of Biscay and the shelf break with depths from 200 to 2000 m, the north Bay of Biscay (which is actually the central section over the shelf break), the south Bay of Biscay where depths can exceed 4000 m and finishes with the complex area of the Iberian Shelf just outside of Bilbao.

Chapter 3

Measurements

The two main data sets used for this study are (i) the level 2 SST_{skin} data from the AATSR sensor on ESA Environmental Satellite (Envisat) and (ii) the *in situ* temperature records collected from ISAR on the PoB and the CpF. The main emphasis in this chapter will be on the *in situ* data because the methodology for retrieving SST estimates from the ISAR records that was described by Donlon *et al.* (2008) was improved by this author and this thesis provides the first systematic assessment of the quality of SST_{skin} data acquired using ISAR. In contrast AATSR data were used as provided by European Space Agency (ESA) (ESA, 2007) and no further investigation of the AATSR retrieval algorithm was made. Thus, the first section of the chapter is a brief overview of the character of SST data acquired from the AATSR. This is followed by a substantial section describing the generic operation and data processing procedures developed to acquire the data from ISAR used in this thesis. Section 3.3 summarises the operational configuration of ISAR for acquiring the data records used in subsequent chapters of this thesis. Section 3.4 specifies the way ISAR was deployed from the PoB, and describes how various contact thermometers were also deployed from PoB to acquire the SST_{depth} data used in Chapter 5. Section 3.5 outlines the arrangement for deploying ISAR on CpF, which had to be implemented when PoB was taken out of service in 2010. Section 3.6 summarises the 8 year SST_{skin} data set acquired from ISAR on both PoB and CpF.

3.1 Satellite SST_{skin}

The satellite SST_{skin} measurements have been acquired with the AATSR instrument (Llewellyn-Jones *et al.*, 2001). This is the third instrument in a series of dual view infra-red radiometers. The AATSR is an along track scanning radiometer, with a swath width of 500 km. Each scan samples first the forward view, angled at 55 degrees from nadir, a hot BB with a temperature of 305 K, a visible calibration target, the nadir view and a cold BB which floats around ambient temperature of the sensor on the satellite which is approximately 256 K. The temperatures of the two BB are designed in such a way that the temperatures are just below and above the expected measurement temperature range. Figure 3.1 shows the unique dual view design of the AATSR instrument which enables an atmospheric correction which can cope with a wider range of atmospheric conditions (Barton *et al.*, 1995) compared to conventional satellite radiometer designs. Apart from the unique dual view design the AATSR has a Stirling Cycle cooler which keeps the detector at 80 K, thus improving the thermal noise characteristics of the detector. Table 3.1 shows the AATSRs 7 measurement channels, three in the visible wave band and four in the infra-red wave band. The AATSR was designed to deliver *SST* with a better than 0.3 K absolute accuracy and a potential stability of less than 0.1 K drift per decade, which is required for climate studies. For an in depth description of the AATSR instrument see Llewellyn-Jones *et al.* (2001).

TABLE 3.1: AATSR channels, from Llewellyn-Jones *et al.* (2001).

Channel	Centre Wavelength	Bandwidth	Primary Application
0.55 μm	0.555 μm	20 nm	Chlorophyll
0.66 μm	0.659 μm	20 nm	Vegetation Index
0.87 μm	0.865 μm	20 nm	Vegetation Index
1.6 μm	1.61 μm	0.3 μm	Cloud Clearing
3.7 μm	3.70 μm	0.3 μm	SST
11 μm	10.85 μm	1.0 μm	SST
12 μm	12.00 μm	1.0 μm	SST

The AATSR N1 data files provide four *SST* products, which are listed in Table 3.2. Further to the four *SST* products the SSES scheme as outlined in Corlett and Poulter (2008) and appendix C has been applied to the N1 data so that comparison could be

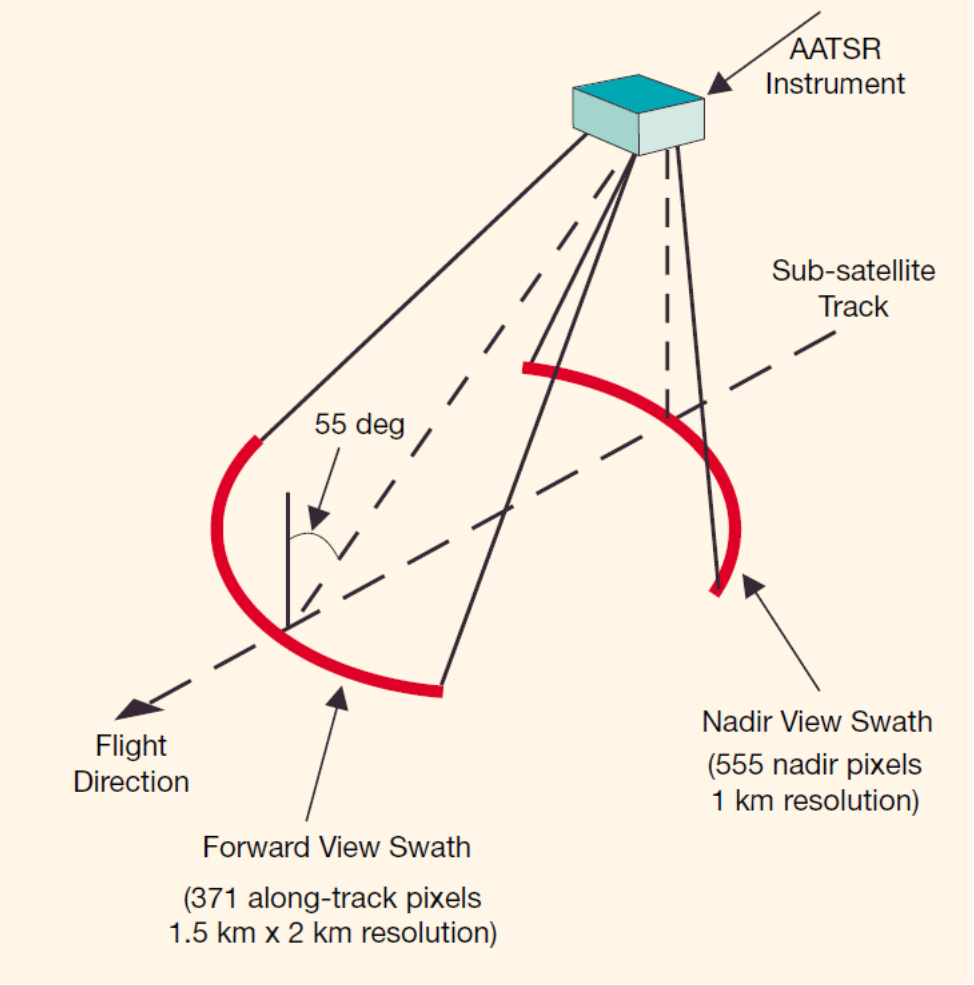


FIGURE 3.1: AATSR view geometry schematics, from Llewellyn-Jones *et al.* (2001)

made between the standard N1 data and the SSES stratified data as described in chapter 6. The reason why this study implemented the SSES scheme on the N1 data rather than using the ESA produced L2p files (GHRSST, 2005) was that the L2p files provide the nadir view data only when there is a dual view present. A nadir view pixel without a coincident forward view pixel and therefore no dual view pixel has generally been taken to imply that there is some cloud contamination in the forward view which makes the nadir view of questionable quality. However, to find an objective measure of whether these nadir-only pixels can be used in high quality data sets, the nadir-only pixel were investigated separately for this study.

TABLE 3.2: List of data products for the level 2 AATSR

SST	Day/Night	Product describtion
N2	Day	Nadir only, two channel (11 μm & 12 μm)
D2	Day	Dual view, two channel (11 μm & 12 μm)
N3	Night	Nadir only, three channel (3.7 μm , 11 μm , 12 μm)
D3	Night	Dual view, three channel (3.7 μm , 11 μm , 12 μm)

3.2 *In situ* radiometric measurements - ISAR

Measuring SST_{skin} from ships is now a fairly mature technology with the ISAR instrument (Donlon *et al.*, 2008) being the first *in situ* radiometer for autonomous use. The ISAR instrument has at the time of writing (2012) produced one of the longest *in situ* SST_{skin} records spanning over 8 years of continuous data acquisition. The ISAR instrument was designed by C Donlon with assistance from G Fisher, I Robinson and R Collins at National Oceanography Centre, Southampton (NOCS), M Reynolds and R Edwards at Brookhaven National Laboratory (BNL) and T Nightingale at Rutherford Appelton Laboratory (RAL). The author of this thesis however has made various improvements to the operating software on the instruments as well as changes to the electronics which made a truly autonomous operation possible. A more detailed description of the instrument can be found in Donlon *et al.* (2008).

3.2.1 Basic design of ISAR

The ISAR instrument contains several subsystems described in outline as follows. (i) The optical system shown schematically in figure 3.2, consists of a scan mirror, a ZnSe window to protect the electronics and the detector from the elements and a Heitronics KT15.85D (KT15) detector. The scan mirror is a 3 mm thick gold surface on copper substrate flat mirror housed in a protective scan drum, which is driven by a DC motor attached to an absolute rotary encoder for accurate positioning. The scan mirror points the field of view successively at the different targets (sea, sky and both BBs, as shown in figure 3.2 (b)), viewed via a small aperture cut into the scan drum wall. This small aperture hole in the scan drum, and subsequently the scan mirror, is the only part of

the optical path which is exposed to the elements. The scan drum is designed in such a way that, when the radiometer target view is outwards the black body cavities are closed and therefore sealed from the elements. Similarly when the scan drum target view points towards one of the BBs, it seals the BBs from the outside as well.

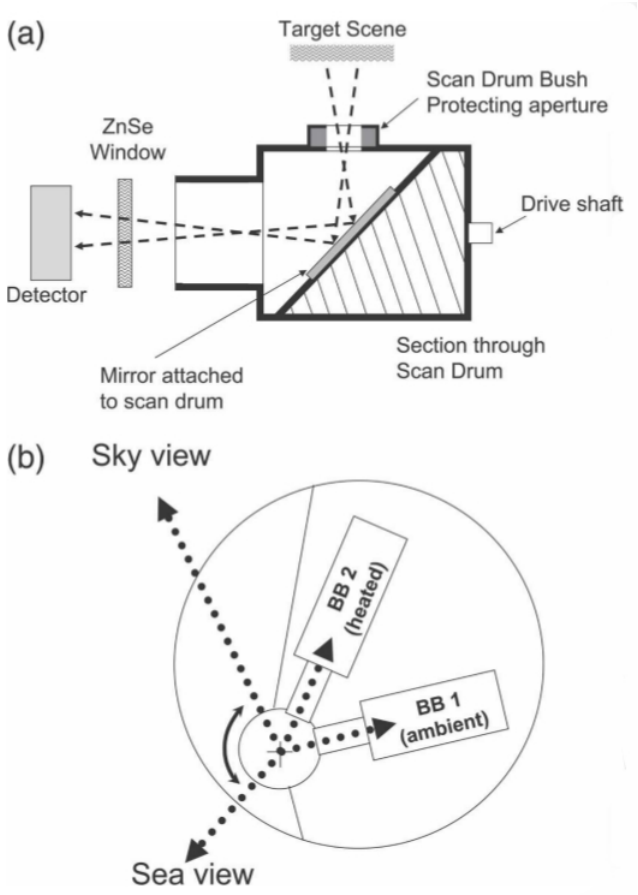


FIGURE 3.2: (a) ISAR optical path showing the main components of the ISAR optical system; the instrument detector (KT15), ZnSe plane Window, scan drum and gold mirror, protective bush (no longer used) and scan drum aperture and calibration BB. (b) location of the ISAR calibration BB cavities in the main instrument body showing the main views made by the ISAR radiometer; sea, sky, BB1, BB2. From Donlon *et al.* (2008)

(ii) The detector and BB calibration system consists of a KT15 which delivers an analogue output dependent on incident radiation within a spectral band-pass of 9.6 - 11.5 μm over a temperature range of 173 - 373 K. The voltage range of the KT15 is set to 0 to 1 V for 173 to 373 K, respectively. During an operating cycle it views sea, sky and two calibration BB housed within the ISAR. The BBs were designed with a re-entrant cone base and a partially closed aperture to have an infra-red emissivity greater than

FIGURE 3.3: External view of the ISAR instrument showing the shutter open. The optical rain sensor and the GPS antennae is pictured in front of the instrument.

0.999. Three precision thermistors (having a NIST traceable calibration to ± 0.05 K) are used to monitor the temperature of each BB.

(iii) The environmental protection subsystem incorporates a storm shutter and an optical rain detector. Whenever the rain detector output rises above a threshold indicating precipitation (or dust particles) the scan drum immediately points inside the ISAR to protect the fore-optics and an external shutter rotates circumferentially to cover the 150° viewing port. When the rain detector output falls below the threshold for a sufficient length of time (normally 10 min) the shutter re-opens and ocean monitoring resumes.

(iv) The internal control and data acquisition system is an on-board computer chip that manages the viewing cycle, controls the shutter operation, performs the analogue-to-digital conversions for the radiometer and thermistor outputs and logs the data. It also monitors a number of other variables logged in the ISAR, such as GPS location and time which uniquely identify every data record, pitch and roll, power supply voltages and the ambient temperature inside the ISAR. It operates the measurement cycle which, for this satellite validation work, is set at 40 samples viewing the sea, 30 samples each viewing the two BBs and 10 samples viewing the sky, a cycle which takes about 240 s to

complete. It uses the BB views to calibrate the detector and then uses the sea and sky views to calculate the skin temperature of the sea, making allowance for sky reflection and the non-blackness of the emission from the sea. It logs the average values for each scan cycle on internal flash memory and sends the full resolution data to an external logging computer. However, it is capable of autonomous operation independently of the external computer.

(v) The external interface uses RS232 protocol to communicate with the external logging computer. The interface also allows for ancillary atmospheric and ocean monitoring sensors mounted on the vessel to be powered and to submit their data into the ISARs internal data logging system. The subsystem components described above are housed in a compact (570 x 220 mm) cylinder shown in Figure 3.3.

3.2.2 A self calibrating radiometer

The baseline internal calibration method must allow for the fact that, although the radiometer field of view is constrained by the field stop in front of the scan drum mirror, it is never possible to eliminate all stray radiation emitted by, or reflected from, inside the radiometer. The calibration of the externally viewed radiances is based on comparing the detector signal when viewing outwards to that when viewing the calculated radiance from an internal BB cavity of known temperature, assumed to fill the same field of view as the external aperture. If it can be assumed that a large proportion, p, of the radiance reaching the detector is from the defined field of view, then the total radiance, L_d, reaching the detector when viewing a target with radiance L_T must be:

$$L_d = pL_T + (1 - p)L_{amb} \tag{3.1}$$

where L_{amb} is the ambient stray radiation inside the sensor. It is further assumed that the instrument signal (that is, its output in counts, C) is proportional to radiance over the range of BT of concern to the ISAR measurements. It is convenient to introduce the variables X_{sea} and X_{sky} which are functions of the target count and the BB counts

recorded during the same ISAR measurement cycle. The target counts C, which are necessary for the definition of X_{sea} and X_{sky} are defined as follows:

$$C = gL_d \tag{3.2}$$

where g is the internal gain of the detector in counts per radiance units. For each measurement cycle the system records the average signal, C_{sea} and C_{sky}, when the scan mirror points the field of view through the external aperture towards the sea or the sky. It also points internally to the ambient BB (BB1) and the heated BB (BB2) recording average counts of C_{bb1} and C_{bb2} respectively. Therefore X_{sea} can be defined as:

$$X_{sea} = \frac{C_{sea} - C_{BB1}}{C_{BB2} - C_{BB1}} \tag{3.3}$$

and X_{sky} as

$$X_{sky} = \frac{C_{sky} - C_{BB1}}{C_{BB2} - C_{BB1}} \tag{3.4}$$

Substitution of 3.2 and 3.1 in 3.3 eliminates the unknowns p, g and L_{amb}, yielding:

$$X_{sea} = \frac{L_{sea} - L_{BB1}}{L_{BB2} - L_{BB1}} \tag{3.5}$$

and hence

$$L_{sea} = X_{sea}L_{BB2} + (1 - X_{sea})L_{BB1} \tag{3.6}$$

A similar equation expresses the sky radiance L_{sky} in terms of X_{sky}. This allows the target radiance to be evaluated for each cycle from a knowledge of the radiances emitted by the two BB, without needing to know p, g and L_{amb}. Moreover, as long as p, g and L_{amb} remain constant within a measurement cycle of one to two minutes, this approach should allow any gradual drift in p, g and L_{amb} to be accommodated without affecting the accuracy of the retrieved target radiance, that is the sea view radiance, the sky radiance or any other target presented to the radiometer such as the laboratory calibration BB. It allows for some degradation of the scan mirror surface which may reduce direct reflection and increase emission by the mirror surface itself so that the proportion $(1 - p)$ of stray

radiation increases. As long as the effect is identical for each view (sea, sky and BBs) during a particular measurement cycle the retrieval of the radiance from external targets using equation 3.6 is not compromised.

The radiances for the internal BBs in equation 3.6 are calculated using:

$$L_{BB} = \varepsilon_{BB} B_B(T_{BB}) + (1 - \varepsilon_{BB}) B_B(T_{amb}) \tag{3.7}$$

where L_{BB} is the radiance leaving the BB in the field of view of the radiometer, ε_{BB} is the effective emissivity of the BB and B_B is the bandwidth adjusted Planck function for the detector bandwidth, evaluated for the BB temperature (T_{bb}) or the ambient temperature (T_{amb}) internally within the ISAR, both of which are measured to a high accuracy by thermistors.

Each ISAR uses a set of polynomial coefficients determined by the KT15 spectral filter data to calculate the bandwidth adjusted Planck and inverse Planck function. These coefficients were determined by Nightingale (2000) and are stored in the instruments configuration file. The KT15 spectral filter has remained stable to date (2012), which has been verified by Combined Action for the Study of the Ocean Thermal Skin second generation black body (CASOTS II) and the radiometer inter-comparisons in 2009 (Theocharous *et al.*, 2010).

3.2.3 Radiometric measurements made by ISAR

Donlon *et al.* (2008) explains the measurement principle of the ISAR along the following lines. The SST_{skin} temperature is estimated by the ISAR instrument by observing both internal BBs, the sky and the sea surface brightness temperatures. The sky view is necessary because the sea surface is not a perfect infra-red radiator. Seawater has an emissivity of slightly less than one and therefore we have to adjust the measured BT of the sea surface to allow for the small amount of reflection of sky radiance which contributes to the sea view measurements. The atmospheric BT is normally cooler than the SST, so the retrieved SST is slightly warmer than the sea view BT (Donlon *et al.*,

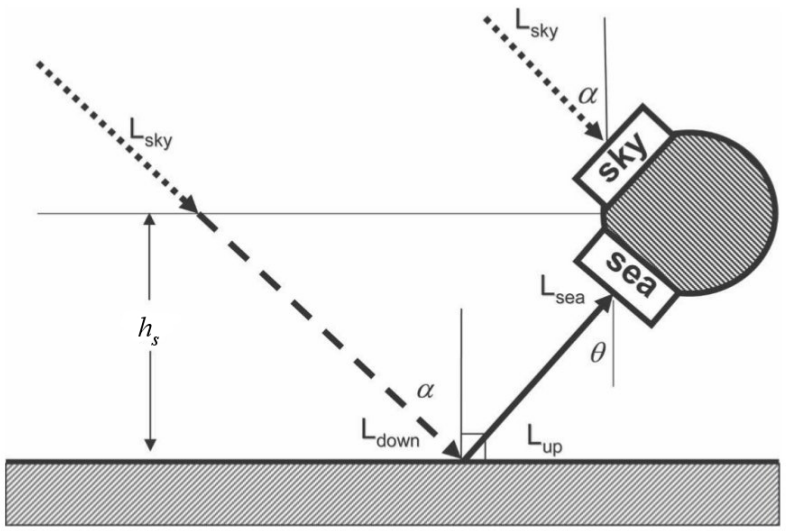


FIGURE 3.4: Geometrical quantities and radiative components that must be considered when measuring the SST_{skin} temperature of the ocean surface. The figure shows a dual port radiometer mounted aboard a ship at a height h above the sea surface viewing the sea surface at a zenith angle θ. From Donlon *et al.* (2008).

2008). The absorption and emission by the atmosphere along the path between the sea and the ISAR must also be considered. Then using the variables defined in figure 3.4 which shows schematically the radiance path reaching the sky and sea views of ISAR, thus:

$$L_{down} \approx \tau_{path} L_{sky}(\lambda) + (1 - \tau_{path}) \overline{B(T_{air}, \lambda)} \tag{3.8}$$

where τ_{path} is the transmittance along the atmospheric path through the atmosphere between the height h_s of the sensor and the sea surface and where $\overline{B(T_{air}, \lambda)}$ is the spectral radiance emitted at T_{air} averaged over the atmospheric path length. It can be evaluated from the Planck function as

$$B(T, \lambda) = \frac{2hc^2}{\lambda^5 \left[e^{\left(\frac{hc}{\lambda k T}\right)} - 1 \right]} \tag{3.9}$$

The upwelling radiance from the sea surface is given by:

$$L_{up} = \varepsilon(\lambda, \theta) B(SST_{skin}, \lambda) + [1 - \varepsilon(\lambda, \theta)] L_{down} \tag{3.10}$$

and the upwelling radiance arriving at the aperture of the radiometer is

$$\begin{aligned} L_{sea}(\lambda) &\approx \tau_{path} L_{up}(\lambda) + (1 - \tau_{path})\overline{B(T_{air}, \lambda)} \\ &= \tau_{path}\varepsilon(\lambda, \theta)B(SST_{skin}, \lambda) + \tau_{path}^2 \\ &\times [1 - \varepsilon(\lambda, \theta)]L_{sky} + (1 - \tau_{path}) \\ &\times [1 - \varepsilon(\lambda, \theta)] + 1\,\overline{B(T_{air}, \lambda)} \end{aligned} \tag{3.11}$$

As τ_{path} approaches unity, L_{sea} is given by

$$L_{sea}(\lambda) = \varepsilon(\lambda, \theta)B(SST_{skin}, \lambda) + [1 - \varepsilon(\lambda, \theta)]L_{sky} \tag{3.12}$$

When the radiometer is less than 40 m above the sea surface and the relative humidity is below $\approx 95\%, \tau_{path}$ is very close to unity for measurements in the region 9.6 - 11.5 μm (Donlon *et al.*, 2008). Donlon *et al.* (2008) states that the working assumption (τ_{path} = 1) in equation 3.11 introduces errors of less than 0.05 K in retrieved values of SST_{skin} and while this error is small, it is not insignificant when the goal is an uncertainty of 0.1 K. The above analysis assumes that the ocean surface is flat and that the component of L_{sky} reflected at the sea surface into the radiometer field of view comes from a zenith angle θ. However, when the sea is rough, radiance from many parts of the sky can be reflected from suitably oriented surface facets into the radiometer field of view (Donlon and Nightingale, 2000). This uncertainty is the subject of ongoing research into the variation of $\varepsilon(\lambda, \theta)$ with sea state (Watts *et al.*, 1996; Wu and Smith, 1997) and to proceed practically Donlon *et al.* (2008) assumed the average reflected sky radiance to be that from the direction that reflects in a calm sea.

The radiometer combined detector spectral response function and wave band filter (if used) is defined as $\zeta(\lambda)$ in output units per unit radiance. The output signal S_{sea} of the radiometer when viewing the sea is then

$$S_{sea} = \int_{\lambda_1}^{\lambda_2} \zeta(\lambda) \left\{\varepsilon(\lambda, \theta)B(SST_{skin}, \lambda) + [1 - \varepsilon(\lambda, \theta)]L_{sky}\right\} d\lambda \tag{3.13}$$

where the limits of integration are chosen to span the bandwidth of the detector and filter as defined by $\zeta(\lambda)$. The output when viewing the sky S_{sky} is

$$S_{sky} = \int_{\lambda_1}^{\lambda_2} \zeta(\lambda) L_{sky} d\lambda \tag{3.14}$$

Within the narrow wave band 9.5 - 11.5 μm , $\varepsilon(\lambda, \theta)$ and $B(\lambda, T)$ vary only slowly with wavelength and so the integral equation 3.13 can be simplified, to a good approximation, into separated form using the band-averaged values ζ_B, $\varepsilon_B(\theta)$, $L_{B,sky}$, and $B_B(T)$, giving

$$S_{sea} = \zeta_B \left\{ \varepsilon_B(\theta) B_B(SST_{skin}) + [1 - \varepsilon_B(\theta)] L_{B,sky} \right\} \tag{3.15}$$

where

$$\zeta_B = \int_{\lambda_1}^{\lambda_2} \zeta(\lambda) d\lambda \tag{3.16}$$

Equation 3.14 can be written

$$S_{sky} = \zeta_B L_{B,sky} \tag{3.17}$$

so that, finally

$$\zeta_B B_B(SST_{skin}) = \frac{S_{sea} - [1 - \varepsilon_B(\theta)] S_{sky}}{\varepsilon_B(\theta)} \tag{3.18}$$

Measurements of the sea and sky radiance responses, S_{sea} and S_{sky}, which are ideally obtained almost simultaneously by looking downward at the incidence angle θ and upward at the zenith angle θ, are required to solve equation 3.18. The time difference between sea and sky measurements must be small to limit uncertainties associated with rapidly changing atmospheric radiance conditions due to clouds of different species and height and therefore different radiative temperature (Donlon and Nightingale, 2000).

Equation 3.18 does not take the self-calibration nature of the ISAR into account, however by combining the radiometric calculation (equation 3.18) and the detector self-calibration (equation 3.6) the SST_{skin} calculation can be written as:

$$\zeta_B SST_{skin} = \frac{B_B^{-1}\left[L_{sea} - [1 - \varepsilon_B(\theta)]L_{sky}\right]}{\varepsilon_B(\theta)} \tag{3.19}$$

where B_B^{-1} is the inverse band averaged Planck function.

Therefore the SST estimation can be summarised with the following steps (i) calibration of the detectors signal for gain and offset by viewing the internal BBs followed by the (ii) calculation of the BB radiances from the BB thermistor temperatures by using the detector filtered band averaged Planck function, then (iii) the BB radiances are corrected for not being emitted from a perfect BB, next the (iv) radiance of the sea and sky can be calculated from the detector signal and the BB radiances, and finally (v) the detector filtered band averaged inverse Planck function is used to convert the sea and sky radiance into temperatures, which together with the $\varepsilon_B(\theta)$ will allow the SST_{skin} to be calculated.

3.2.4 Revised radiometric calculation[1]

When ISAR had been operated for a number of deployments, each of several months duration, it became apparent that the uncertainties when compared against a laboratory BB at the end of a deployment were sometimes larger than the 0.1 K specification. This required a reconsideration of how the ISAR temperature retrieval algorithm copes with progressive degradation of the optical components, and led to changes being made to the calibration algorithm that converts raw counts from the instrument into SST retrievals.

While the laboratory calibration tests before and after each deployment confirmed the general robustness of this approach, it also showed that the degradation occurring by the end of a three month deployment was leading to marginal performance within the $\pm$ 0.1 K specification of the ISAR. The main source of the problem appears to be in

[1]The work presented here has been previously reported in Wimmer and Robinson (2008b) and Wimmer *et al.* (2012).

estimating the absolute radiances leaving each of the black bodies and reaching the scan drum aperture.

These are needed to evaluate equation 3.6, and are calculated as described in equation 3.7. ε_{BB} is a measure of the non-blackness of the BB integrated across the field of view of the radiometer through the scan-drum aperture. This may encounter the edge of the BB aperture, even when the scan mirror is in a pristine state at the start of a deployment. Thus the heated BB, when viewed through the field stop, is probably not perfectly uniform from all directions within the field of view defined by the field stop which is the scan drum aperture. It may have a slightly cooler annular region around its circumference. To allow for this small effect, the emissivity was estimated in the original algorithm to be 0.9993, and the use of this value is confirmed by the pre-deployment laboratory calibrations.

When the scan mirror degrades, not only does it reduce the reflected signal as discussed above, but it is also likely to cause some scattering of the reflected radiation. This means that the radiometer now receives some radiation scattered into its field of view from a wider range of angles beyond the mirror than it does when the mirror is pristine. The surfaces inside the ISAR that determine the magnitude of this scattered radiation are different for different settings of the scan drum, that is the extra stray radiation is different for the external view and each of the BB views. Since the additional radiation will be mostly at the ISARs ambient temperature, the effect is accommodated largely by a further change in p, but this does not allow for any small changes between different scan drum pointing directions. In particular, when viewing the heated BB, the integrated radiance from the new spread of directions associated with the scattering surface is likely to be reduced because it includes mainly the inside of the scan drum housing. To allow for this factor and prevent it from degrading the radiance calibration requires a small reduction of the apparent emissivity of the BBs ε in equation 3.7. Note that the actual emissivity of the internal BBs does not change, however the change in emissivity of the internal BB in equation 3.7 is a practical way to account for the increased stray radiance coming from a degraded scan mirror.

Trial and error showed that reducing ε by a small amount readily achieved a very good match to the post-deployment laboratory calibration, even when the mirror was badly degraded. That discovery by itself could not provide a practical adjustment to ε during a deployment. What is required is to define $\varepsilon_{BB} = \varepsilon_0 - \varepsilon^{'}$ where ε_0 is the pre-deployment effective emissivity set at 0.9993, and then to find a means to vary $\varepsilon^{'}$ continuously as an objective response to the actual degradation of the mirror. Fortunately the mirror response is readily monitored by evaluating a nominal gain factor, G, where:

$$G = \frac{C_{BB2} - C_{BB1}}{L_{BB2} - L_{BB1}} \tag{3.20}$$

and the BB radiances are evaluated using the initial emissivity value $\varepsilon_0 = 0.9993$. That is:

$$L_{BB0} = \varepsilon_0 B_B(T_{BB}) + (1 - \varepsilon_0) B_B(T_{amb}) \tag{3.21}$$

As the mirror degrades during a deployment, p increases so less of the radiance emitted by the hot BB reaches the detector (see equation 3.1), so that (from equation 3.2) C_{BB2} is reduced and thus G decreases as a direct measure of the mirror degradation. At the same time, the use of $\varepsilon = \varepsilon_0$ in (3.7) means the black body emission is slightly overestimated when the mirror degrades, which further reduces G. Thus the difference between G_0 (the value of G evaluated from the pre-deployment calibration) and the subsequent value of G at every measurement cycle of the deployment can serve as an indicator of $\varepsilon^{'}$. A simple linear relationship, $\varepsilon^{'} = f_W(G_0 - G)$, was found to be applicable, with a weighting factor, f_W, of $0.2 - 0.45$. Thus

$$\varepsilon_{BB} = \varepsilon_0 - f_W(G_0 - G) \tag{3.22}$$

The new radiance retrieval algorithm evaluates equation 3.21 and then equation 3.20, in order to obtain ε_{BB} from equation 3.22 at every measurement cycle. The evaluation of equation 3.3 and 3.4 is unchanged from the original algorithm, but then the radiance is derived using the result of equation 3.22 in 3.7 to calculate the correct BB radiance to enter in equation 3.6 which delivers the desired target radiance. Also equation 3.7 was

modified slightly by using measurements of the temperature of the flange surrounding the black body cavity in place of T_{amb} in the second term on the r.h.s. The experience of applying the methods to most of the deployments so far showed that the selection of $f_W = 0.2 - 0.45$ was applicable to most of the deployments. The new algorithm always led to the post deployment laboratory calibration being almost identical to the pre-deployment calibration. This is illustrated in section 3.2.5.2 by an example which compares the laboratory validation of ISAR temperature retrievals using a laboratory infra-red source, with and without applying the modifications to the processing algorithm.

Note that G which is readily calculated from the output records from each scan cycle, can therefore also be used as a diagnostic decision tool for aborting a deployment earlier than planned if a sudden major mirror degradation is noted.

3.2.5 Quality control of radiometric measurements

To ensure that the ISAR measurements are of the highest quality, within the quoted accuracy of ± 0.1 K and traceable to SI standards a number of procedures are followed as listed below:

1. Checking that the instrument performs within its design parameters, by examining the RAW data.
2. Verifying that the instrument measures with an accuracy of ± 0.1 K by using an external BB target.
3. Ensure that the instrument and the external BB are traceable to international standards.

These procedures are elaborated in the subsections which follow.

3.2.5.1 ISAR engineering data

As an initial step to verify whether the ISAR instrument is working satisfactorily a set of so called engineering data plots is produced from the raw ISAR data file. These

plots show all the key raw instrument data, such as voltages, the derived data, such as temperatures, together with position and roll and pitch information. A description of the ISAR data fields can be found in the ISAR user manual (Donlon and Wimmer, 2007). Figures 3.5 to 3.14 show the the engineering data plots for a data file from deployment 22 on the PoB between 10.12.2008 and 06.01.2009.

The **geographical location** where the measurements were acquired are shown in the top panel of figure 3.5. The lower panel in figure 3.5 shows the **scan drum position** (in blue) together with the **shutter position** (in red). The green diamonds on the bottom panel show that the shutter end state (open or closed) was reached on the the open/close movement. It is important to know that the shutter is fully open or closed, as a partially open shutter can obstruct the field of view and would therefore give inaccurate SST_{skin} measurements. The scan positions represent at 280° the ambient BB, at 325° the heated BB, at 25° the sky view and at 155° the sea view. The left hand scale of the bottom panel shows the scan position in degrees and the right hand scale the shutter position, where 0 represents a closed shutter and 1 an open shutter. Gaps in the sea and sky view scan drum angles are due to shutter closures in response to rain or sea spray.

The **thermistor temperatures** as recorded by the ISAR instrument are shown in figure 3.6. The top panel shows the temperatures recorded by thermistors attached to various components in order to monitor the internal thermal environment of the ISAR. These are the temperature of BB one (BB1, ambient BB, in blue), the ZnSe window temperature (red), the KT15 case temperature (magenta), the micro processor electronics temperature (green) and PNI roll and pitch sensor (PNI) temperature readout (yellow). The lower two panels show the temperatures recorded by higher precision thermistors which are embedded in the two BB cavities. These are used to characterise the BB radiation emitted by the cavities. Their performance is therefore influential in the overall accuracy of temperature retrieval from the ISAR. In the middle panel the temperatures of the BB 1 (also known as the ambient BB) are shown, the temperature of the first base thermistor (thermistor in the middle of the re-entrant cone) in blue, the temperature of the second base thermistor (thermistor located halfway between the center of the re-entrant cone and the side wall) in red and the temperature of the

aperture thermistor in magenta. In the bottom panel the temperatures of the BB 2 also known as the heated BB are shown, the labelling is the same as for BB 1. The BB thermistors in the BB and the window thermistor are calibrated to an accuracy of $\pm 0.05K$ (Measurement Specialties, 2008). The KT15 thermistor is calibrated to an accuracy of $\pm 0.2K$ (Measurement Specialties, 2008) and the micro processor electronics temperature and PNI temperature readout are not calibrated.

Figure 3.7 shows the **thermistor voltages** as recorded by the on-board Analogue Digital Converter (ADC) for the thermistors as described for figure 3.6. The PNI thermistor has a digital readout and is therefore missing from this plot.

In figure 3.8 the **KT15 analogue output** is shown. The top panel shows the analogue detector signal where the colour represents the different target views. In blue the detector signal of the ambient BB is shown, in red the detector signal of the heated BB, in green the detector signal of the sky view and in magenta the detector signal of the sea view. The bottom panel shows the BT of the four target view as calculated by the KT15 with an emissivity of 1. Gaps in the sea and sky view data are due to shutter closures because of rain or sea spray.

The **GPS and the roll and pitch** data are shown in figure 3.9. The top three panels show the GPS output for the speed over ground, the course made good and Magnetic variation (VAR). The bottom three panels show the PNI output for the roll, the pitch and the magnetic heading of the ship. The magnetic compass is not calibrated for the ship so the heading information is not accurate and should be ignored.

Figure 3.10 shows the input **power supply voltage** in the top panel. The second panel from the top shows the **reference voltage supply** output. The reference voltage supply is the supply for the thermistor half bridges and therefore a critical component in the quality assurance of the ISAR instrument. The third panel from the top shows the estimated SST_{skin} calculated as per equation 3.18, again data gaps are due to shutter closures because of rain or sea spray. The fourth panel from the top shows the shutter state, one for an open shutter and zero for a closed shutter and the bottom panel shows the rain gauge signal in Volts.

Figure 3.11 is a zoom out of the bottom panel of figure 3.10 with the **rain gauge signal** in blue (left hand scale) and the shutter state in red (right hand scale). The purpose of this plot is to check the functionality of the shutter and to make sure it triggers at the correct rain gauge signal.

The **BT** of the target views as estimated by the *SST* processor are shown in figure 3.12. The BT are the inputs for equation 3.18. The top panel shows the BT of the sky view and the bottom panel shows the BT of the sea view in blue and the estimated *SST* in magenta.

Figure 3.13 shows the **mirror gain factor**, which is calculated as described in equation 3.20. The top panel shows the mirror gain factor as a voltage to temperature ratio and the bottom panel shows the mirror gain factor as voltage to radiance ratio. In red is the actual data in both panels and in blue a 100 point averaging window.

In figure 3.14 the source data for figure 3.13 is shown. In the top panel the ZnSe window thermistors temperature is shown in blue when the scan drum points to BB 2 and in red when the scan drum points to BB 1. The second panel shows the temperature difference between BB 2 and BB 1. The third panel shows the signal difference of the KT15 between BB 2 and BB 1 and the bottom panel shows the radiance difference between BB 2 and BB 1 as calculated from the KT15 signal. In the bottom three panels the red dots represent the actual data and the blue dots a 100 point averaging window through the data.

The diagnostic plots are used to verify that the instrument was operating within its design parameters during a deployment. The plots are also used in case of an anomalous output temperature to identify any technical issues, which could cause such an anomalous temperature.

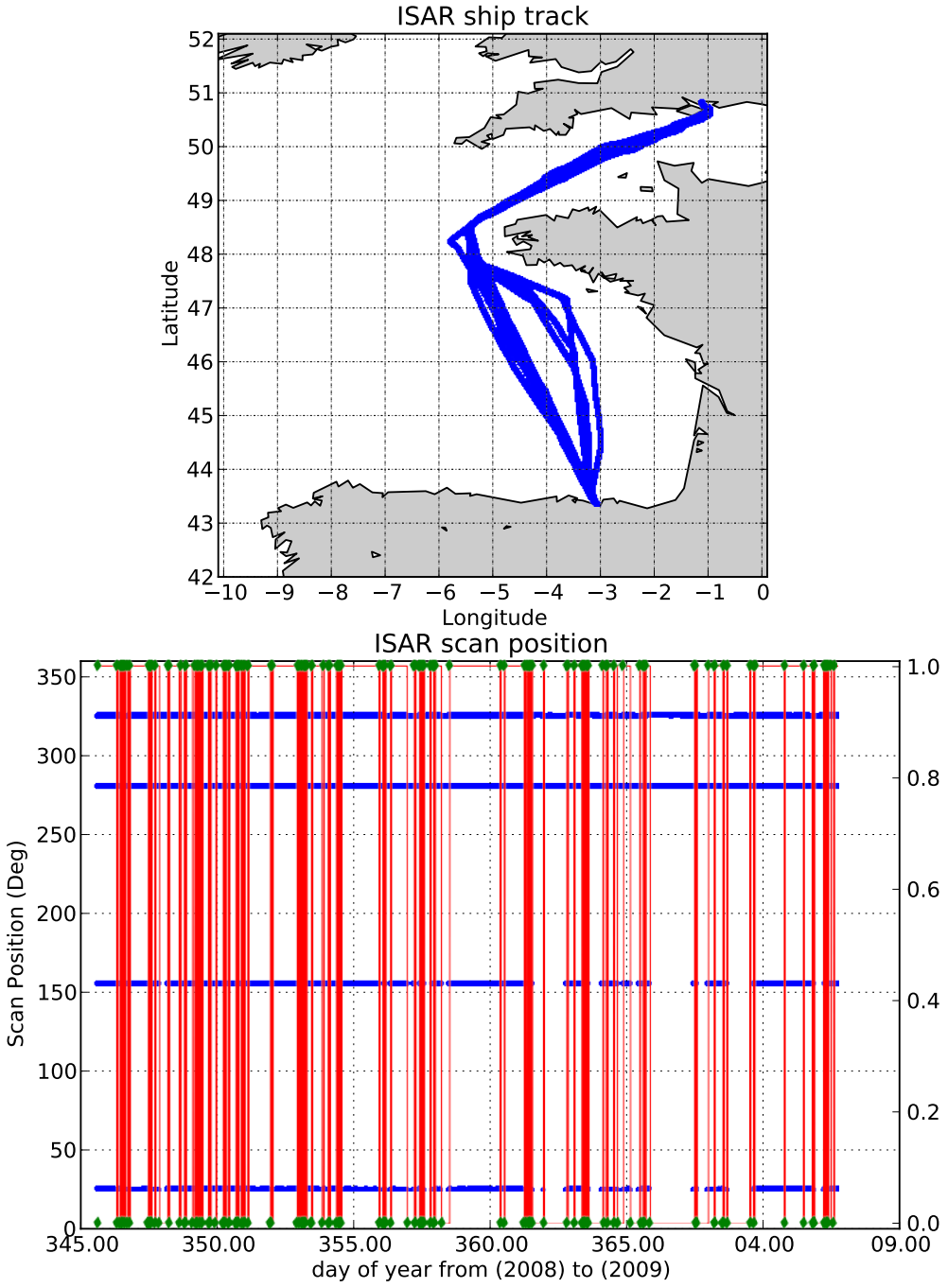


FIGURE 3.5: The ISAR engineering plot for D22, the image shows the ship track in the top panel and scan drum position (in blue, left hand scale) with the shutter state (red, right hand scale) in the bottom panel.The green diamonds show that the shutter has reached its end position.

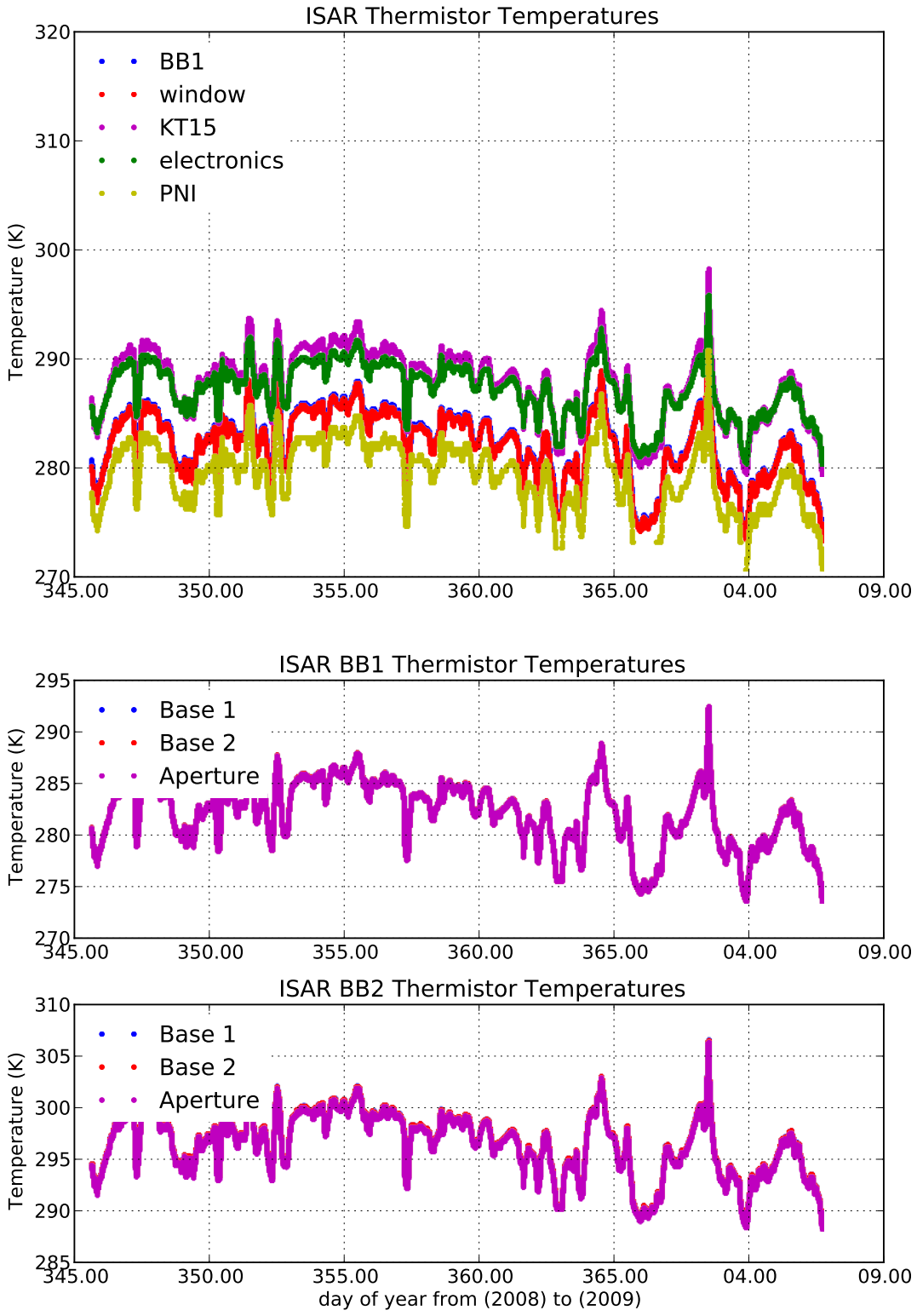


FIGURE 3.6: The ISAR engineering plot for D22, the image shows the instrument thermistor temperatures; in the top panel various thermistors that are distributed throughout the instrument are shown, in the middle panel the ambient temperature black body thermistors temperatures are shown and in the bottom panel the heated temperature black body thermistors temperatures are shown

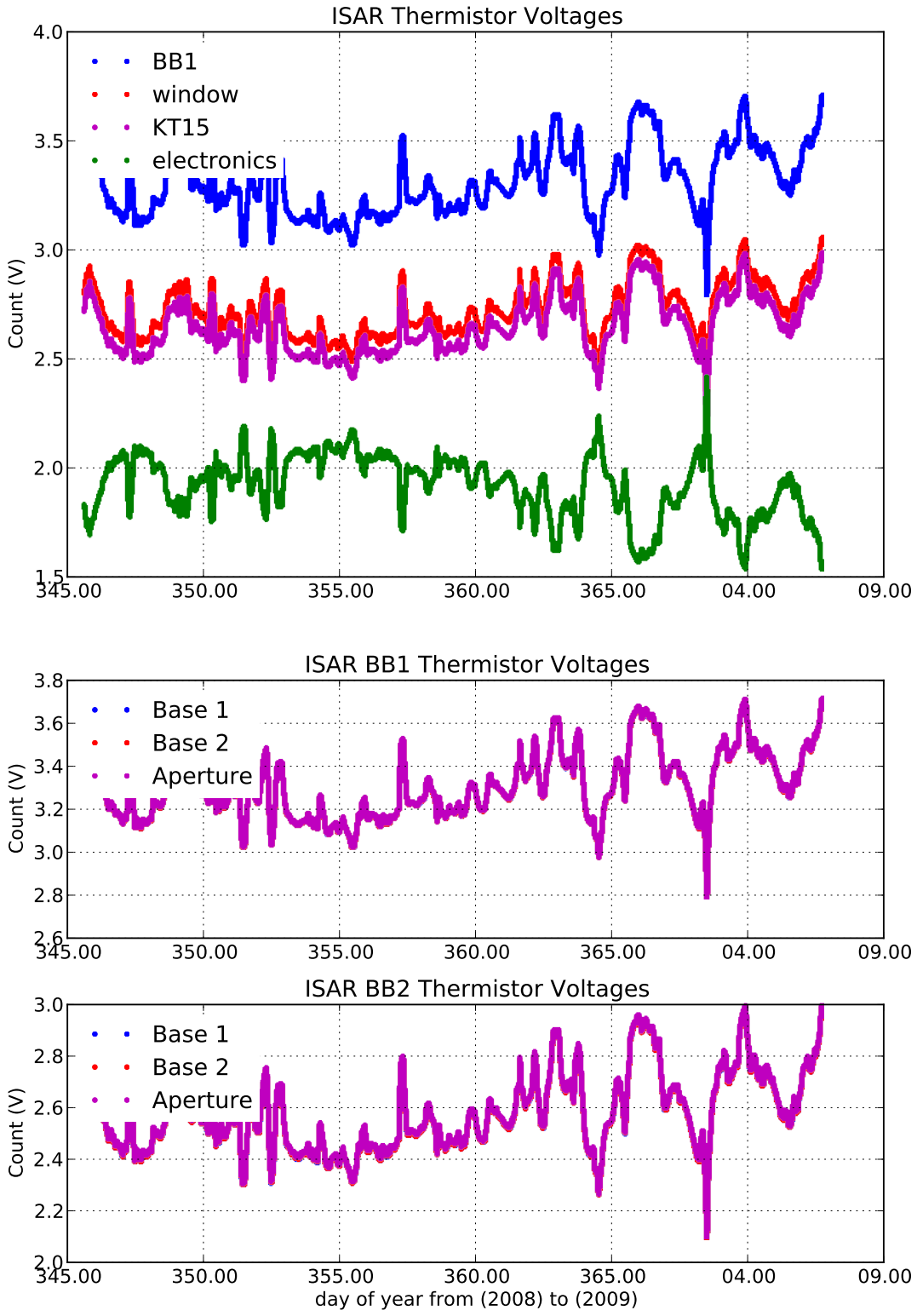


FIGURE 3.7: The ISAR engineering plot for D22, the image shows the instrument thermistor voltages; in the top panel various thermistors which are distributed throughout the instrument are shown, in the middle panel the ambient temperature black body thermistors voltages are shown and in the bottom panel the heated temperature black body thermistors voltages are shown

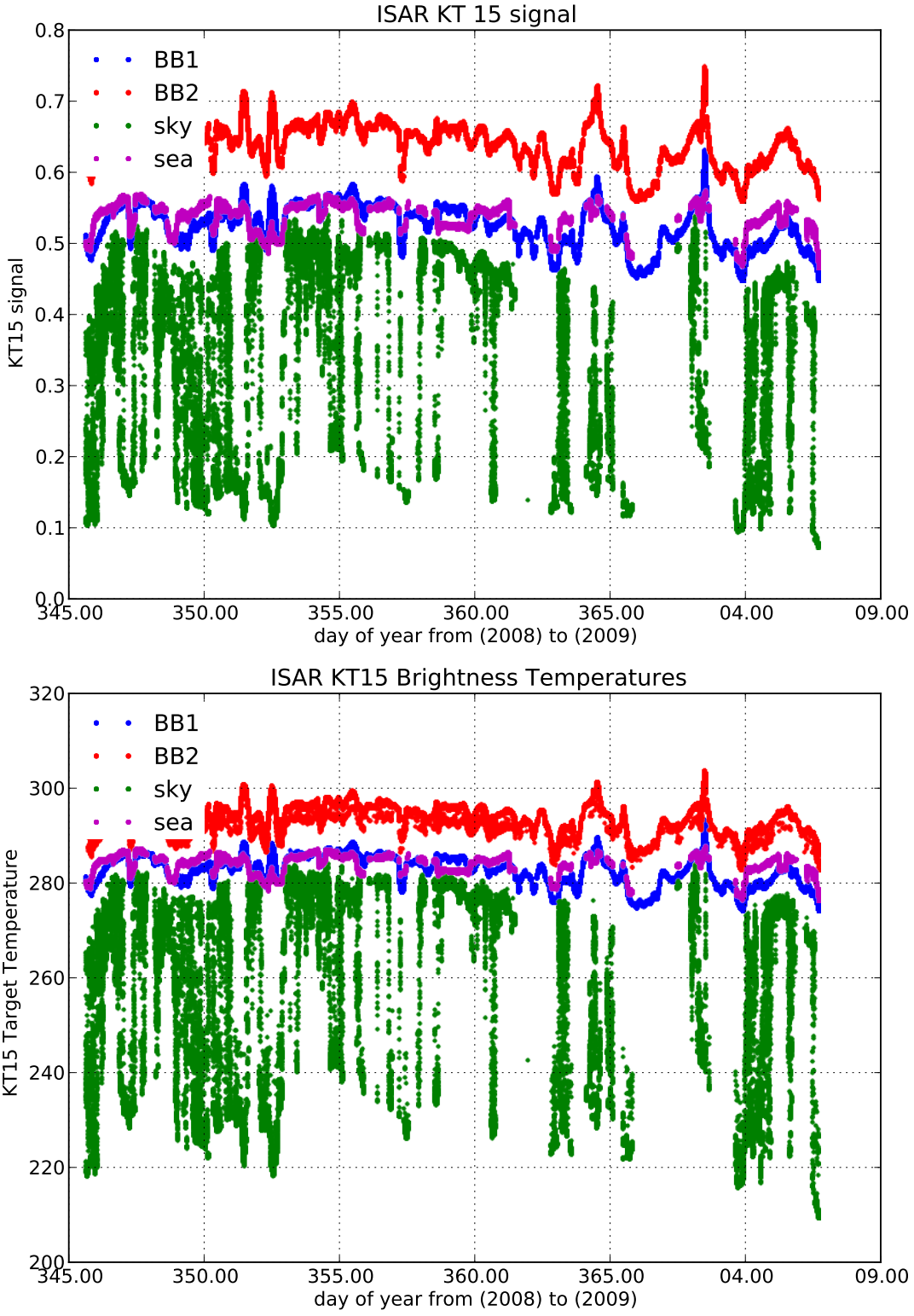


FIGURE 3.8: The ISAR engineering plot for D22, showing the detectors response in the top panel as voltages and in the bottom panel as brightness temperatures calculated with an emissivity of 1.

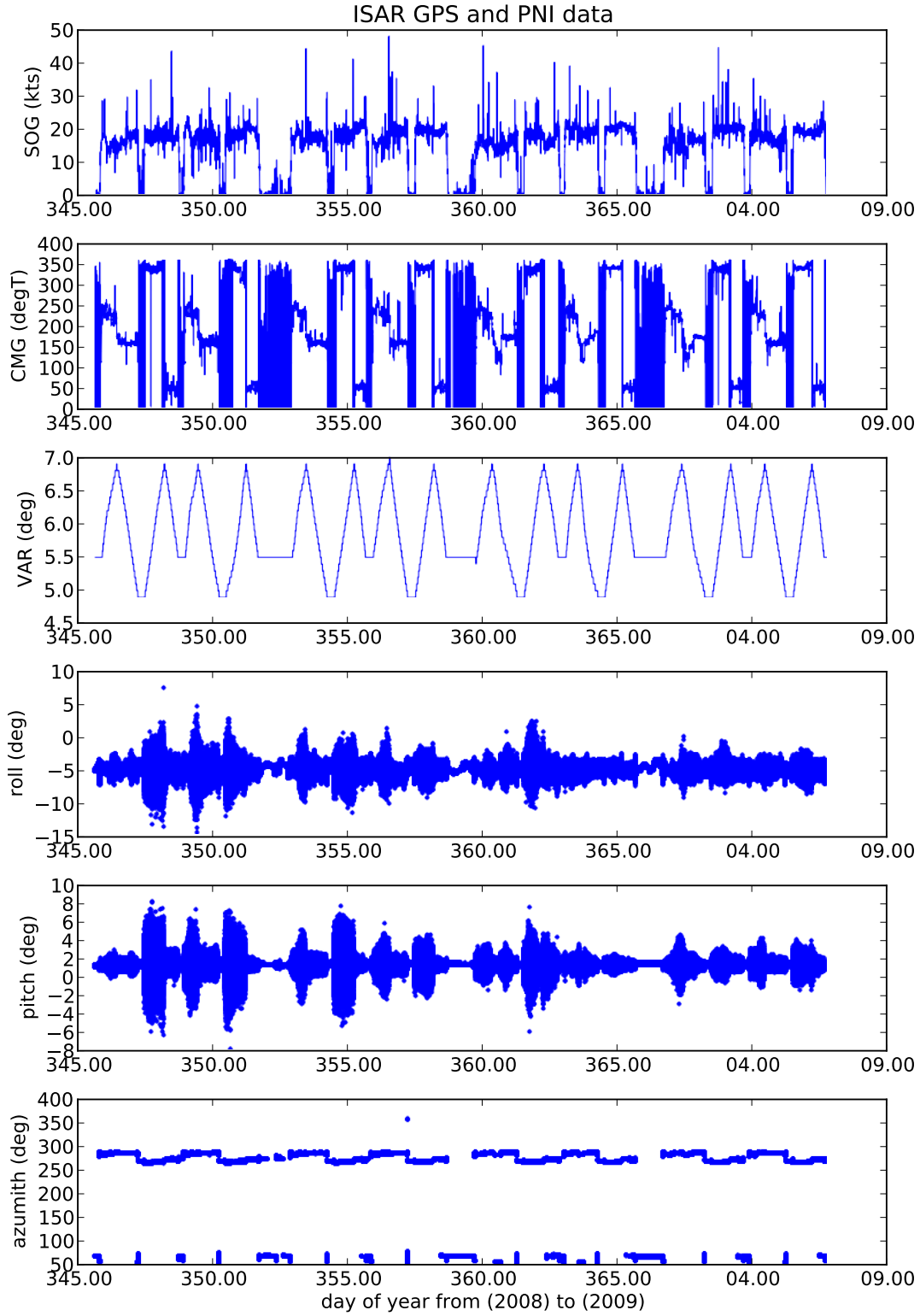


FIGURE 3.9: The ISAR engineering plot for D22, shows the GPS module extra outputs - SOG, CMG, VAR - and the PNI module output in the bottom three panels.

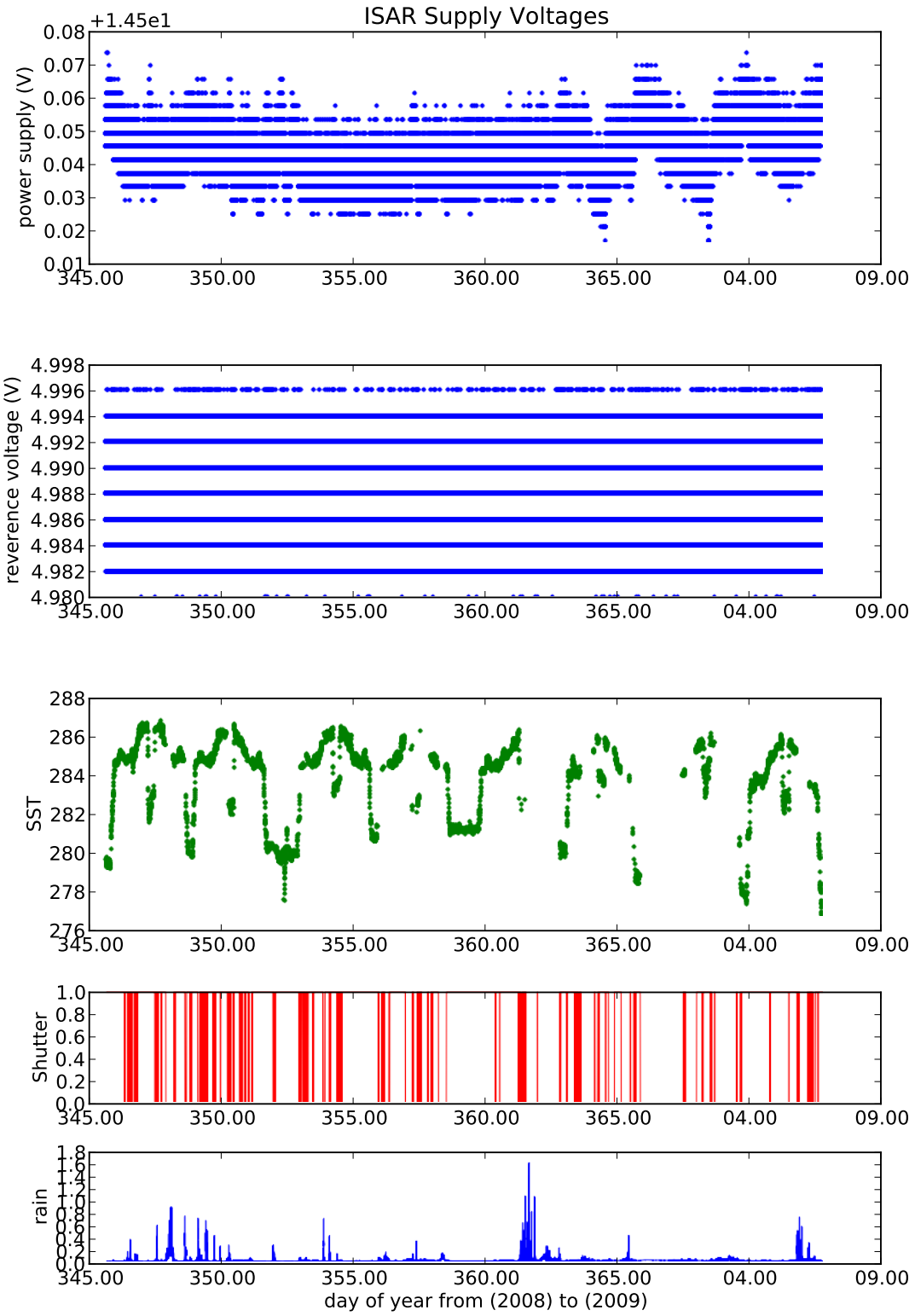


FIGURE 3.10: The ISAR engineering plot for D22, shows the input power voltage to the instrument in the top panel, the reference voltage used for the black body thermistors in the second panel, the *SST* in the third panel, the shutter state in the fourth panel and the rain gauge signal in the fifth panel

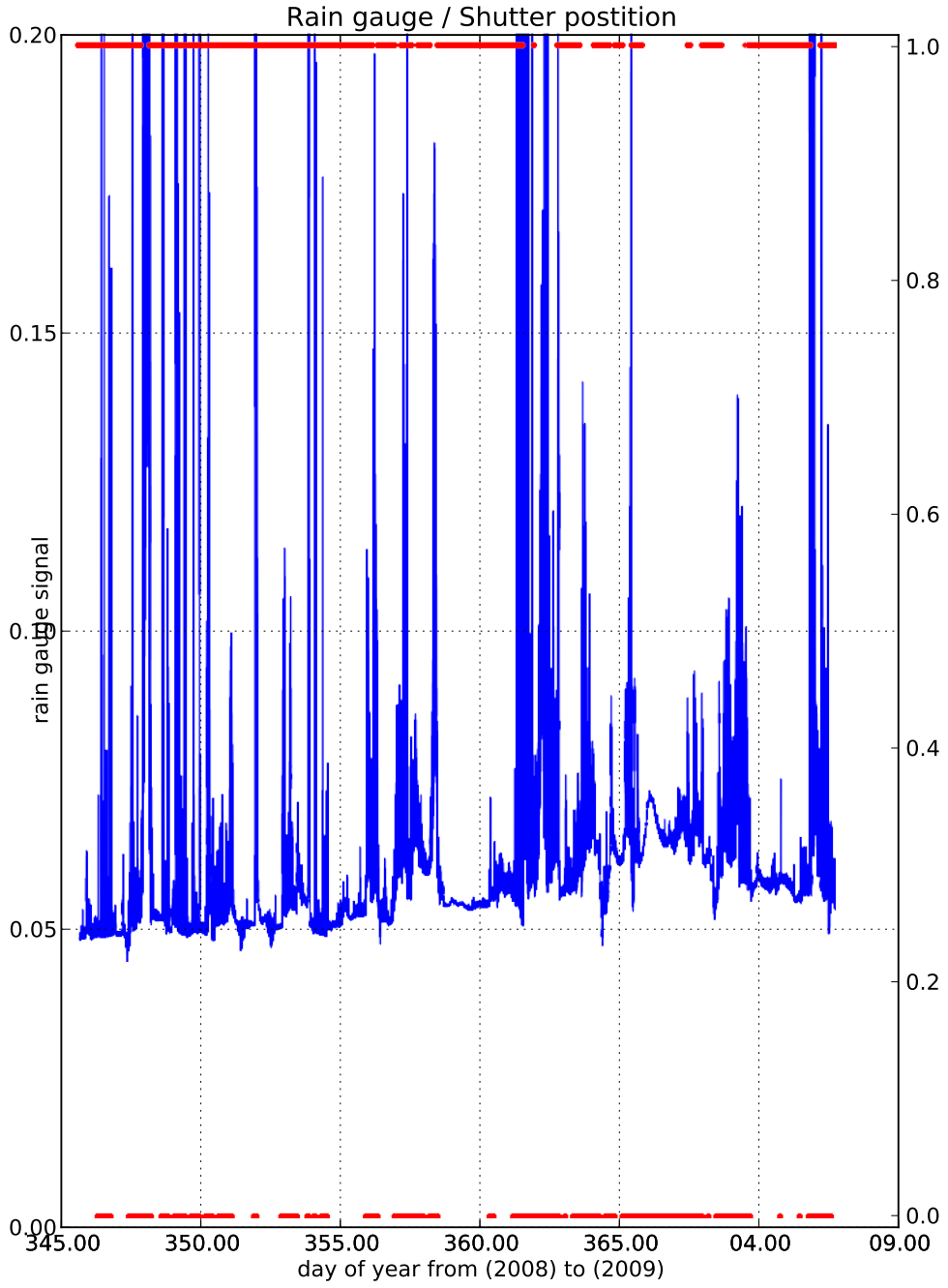


FIGURE 3.11: The ISAR engineering plot for D22, shows a zoomed section of the rain gauge signal (blue) with the shutter state superimposed in red.

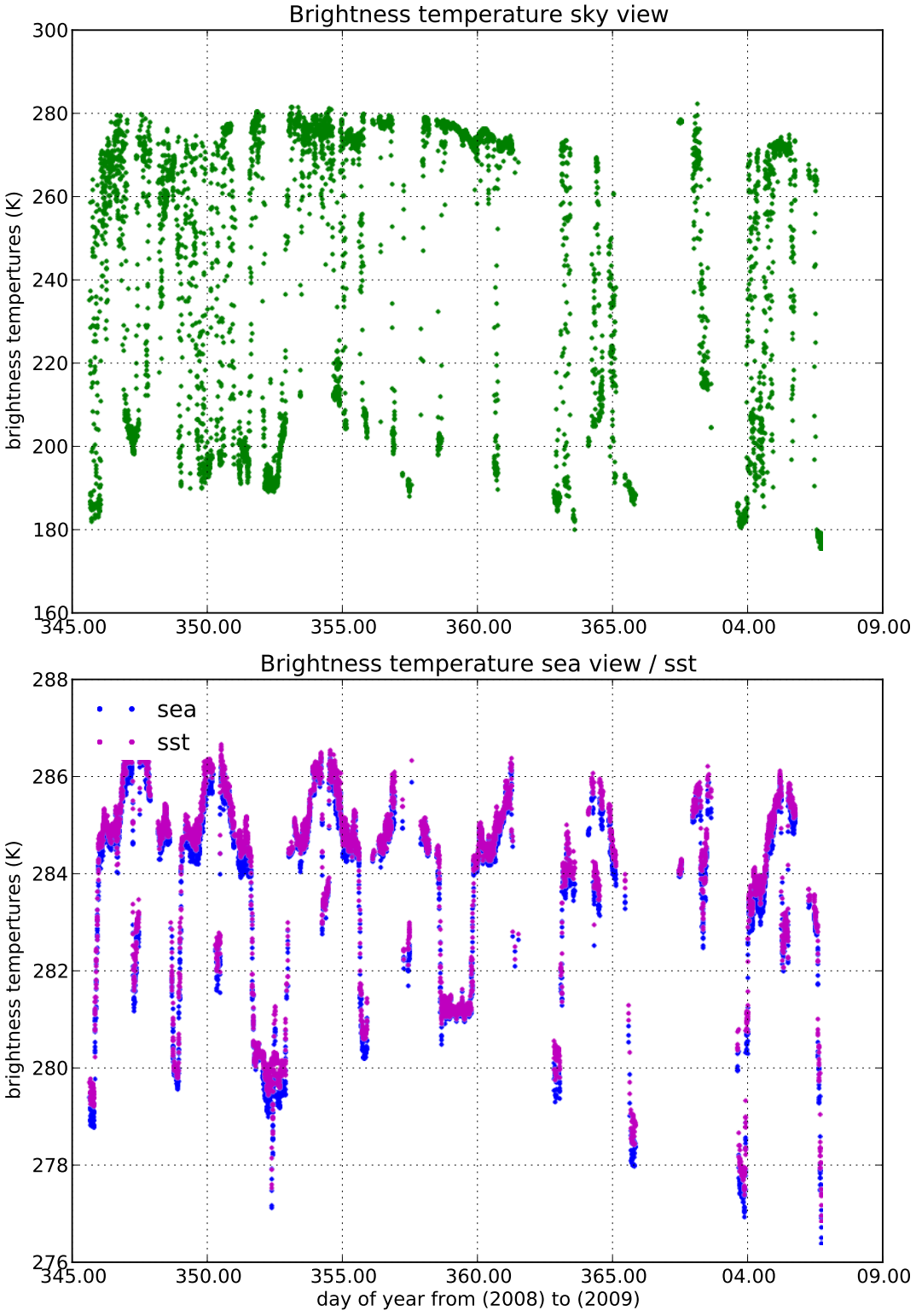


FIGURE 3.12: The ISAR engineering plot for D22, shows the sky brightness temperature in the top panel and the sea brightness temperature (in blue) and the SST_{skin} (in magenta) in the bottom panel.

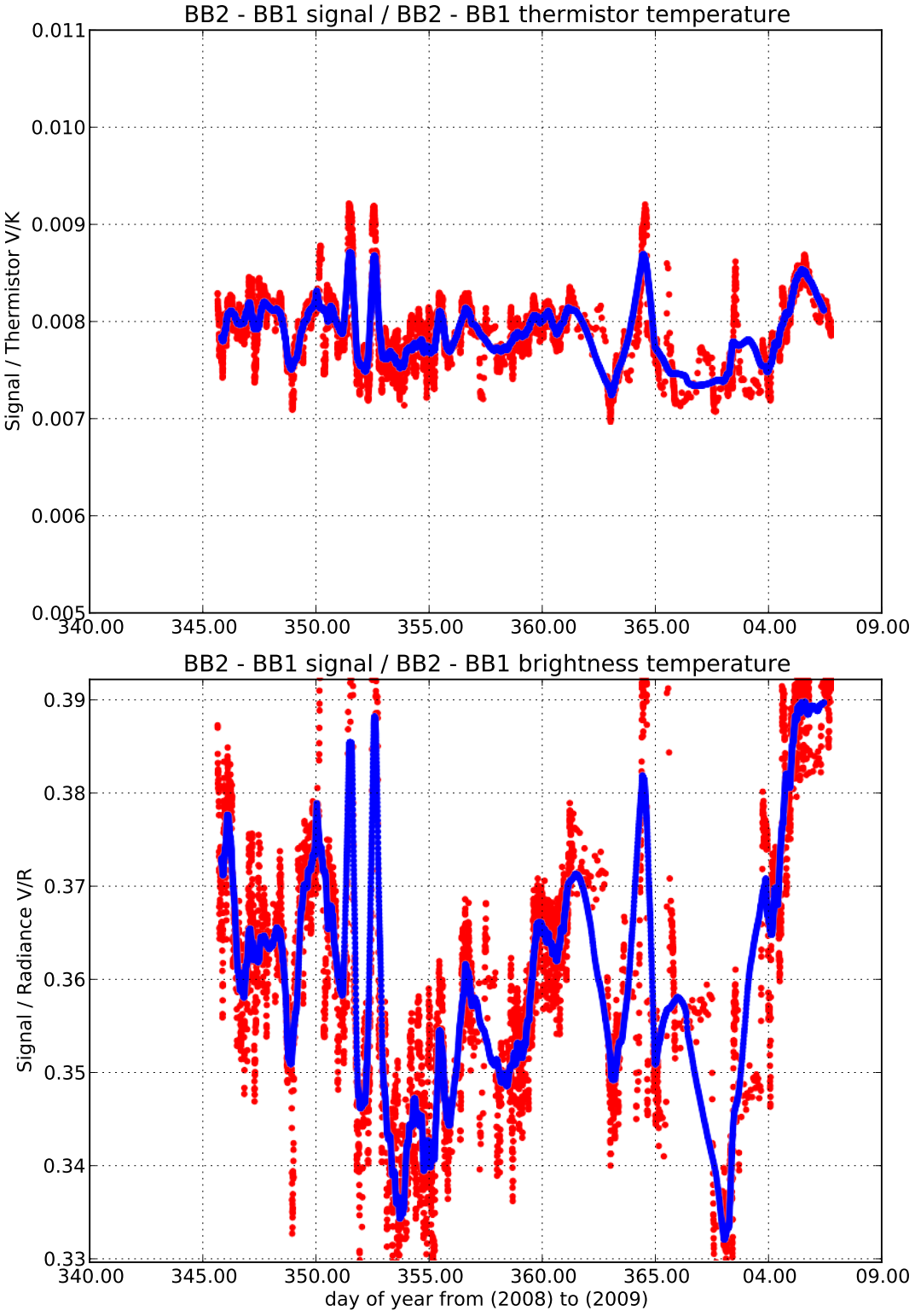


FIGURE 3.13: The ISAR engineering plot for D22, shows the gain factor Equation 3.20 calculated with temperatures in the top panel and radiances in the bottom panel. The blue line is a 100 point moving window average.

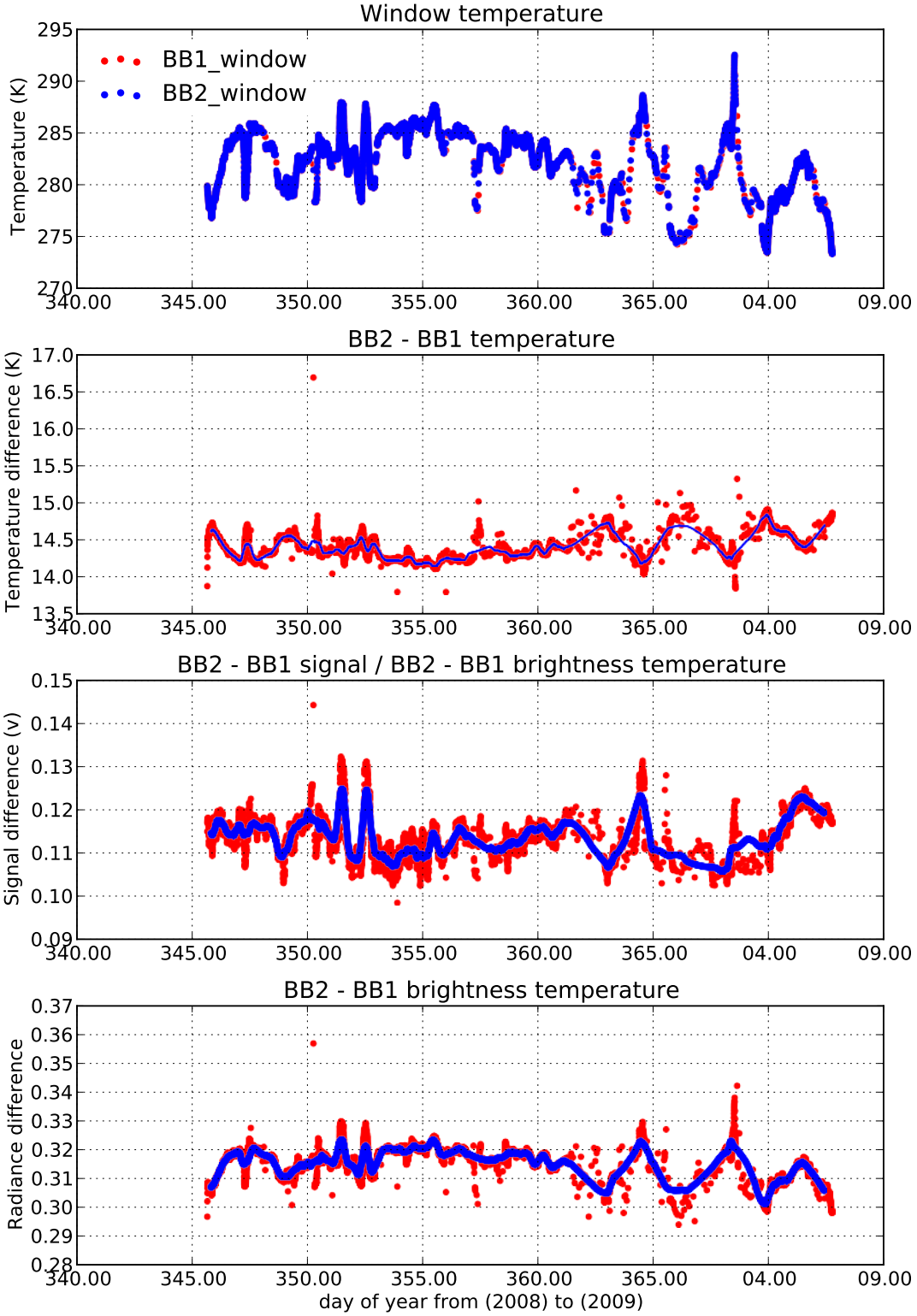


FIGURE 3.14: The ISAR engineering plot for D22, shows the window temperature split with viewing angle in the top panel, the temperature difference between the ambient and heated black body as temperature difference in the second panel, as voltage difference in the third panel and as radiance difference in the bottom panel. The blue line in the bottom three panels is 100 point moving window average.

3.2.5.2 Verification against a black body source

The ISAR instrument's quality control is paramount to ensure the highest possible data quality. This is achieved not only by the instrument's internal BBs but also by using a laboratory BB. The piece of equipment used for this is called the CASOTS II , which is a modified form of a BB produced as the outcome of a 1990's project called Combined Action for the Study of the Ocean Thermal Skin.

Figure 3.15 shows the CASOTS II black body with the infra-red radiometer calibration system. The CASOTS II (Donlon *et al.*, 2013) is an improved version of the Combined Action for the Study of the Ocean Thermal Skin (CASOTS) design (Donlon *et al.*, 1999) which was based on a third generation Geist and Fowler black body design (Fowler, 1995). The CASOTS II water bath is stirred by a powerful pond pump (4000 l/h) and the water bath temperature is measured by a HART 1504 platinum resistance thermometer with a Thermometrics 225 probe.

The ISAR instrument's *SST* is validated against the water bath temperature before and after each deployment. Figure 3.16 shows such a validation before deployment 22 (06.11.2008 to 06.01.2009) and confirms that ISAR was operating within its designed accuracy of $\pm$ 0.1 K. Figure 3.17 shows the initial validation after the end of deployment 22, based on the ISAR *SST* retrieval algorithm before the modification discussed in section 3.2.4 was applied, which was not satisfactory. However an inspection of the instrument and the engineering data, in particular the mirror gain (Equation 3.20 and figure 3.13), showed that the contamination of the mirror was the source of unsatisfactory verification and on application of the revised radiometric calculation (Section 3.2.4) the *SST* retrieval algorithm can be seen to be operating once again within its design parameters (Figure 3.18). In a case where either the verification before or after a deployment is not within the target accuracy of $\pm$ 0.1 K (using the revised *SST* retrieval procedure), no *SST* data are released for further application until the problem is fully understood and (if it is possible to do so in using an objectively justified procedure) corrected.

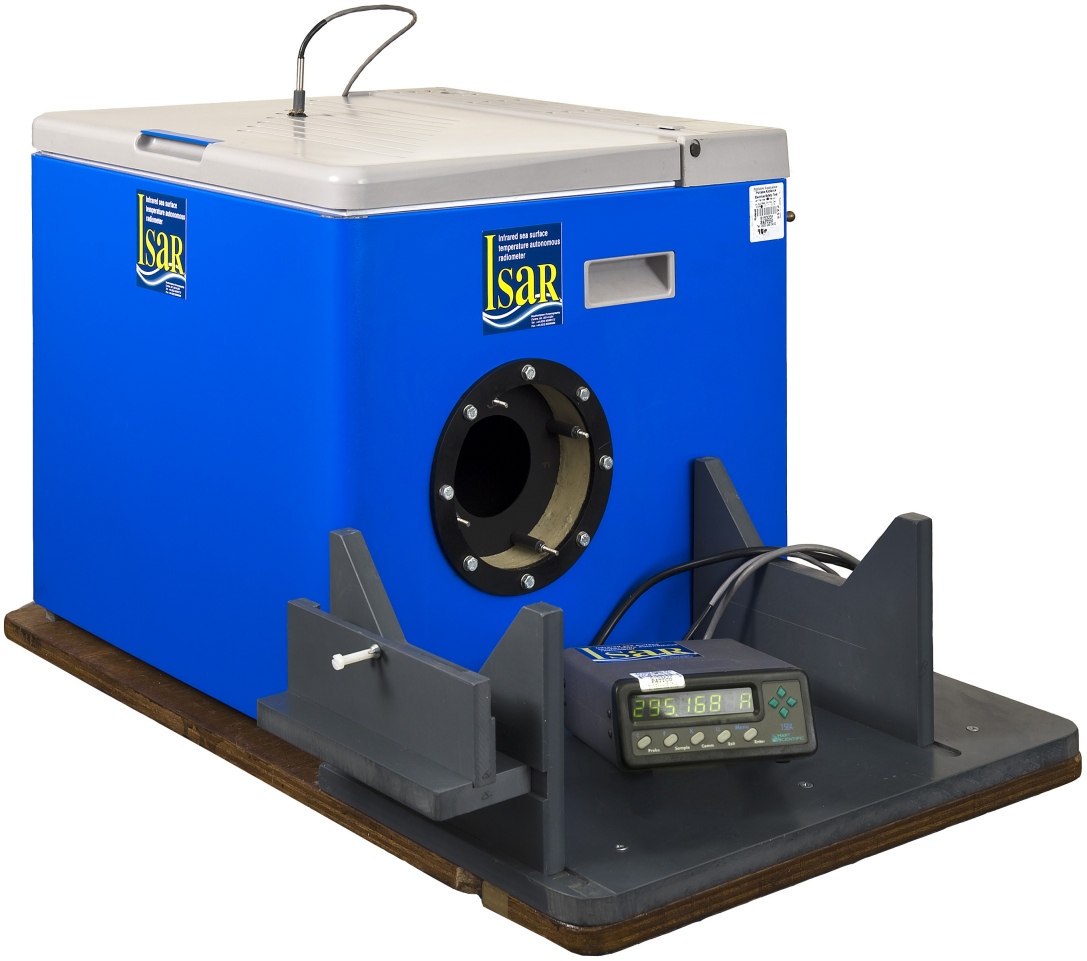


FIGURE 3.15: The CASOTS II black body with the calibration cradle for the ISAR.

3.2.5.3 Traceability to international standards

Because there is no absolute standard for radiometric measurements all measurements have to be traceable to international or national standards so that all the uncertainties can be attributed to the right source. This is especially important with validation projects as the satellite sensor and the *in situ* measurement have to be traceable to the same standard so that the uncertainties can be classified correctly.

The CASOTS II has been characterised against the Rosenstiel School of Marine and Atmospheric Science, University of Miami (RSMAS) NIST black body in 2006 and 2009 and against the NPL black body in 2009, see Theocharous *et al.* (2010) and Theocharous

and Fox (2010). A short excerpt of the results from the 2009 inter-comparison is shown in table 3.3 and 3.4.

TABLE 3.3: ISAR traceability results 2009 (Theocharous *et al.*, 2010). Positive values mean ISAR is reading higher temperatures than the standard.

Temperature	Difference at NPL	Difference at RSMAS
$10°C$	69 mK	below dew point
$20°C$	38 mK	21 mK
$30°C$	6 mK	12 mK

TABLE 3.4: CASOTS II traceability results 2009 (Theocharous and Fox, 2010). Negative values mean CASOTS II is reading lower temperatures than the standard.

Temperature	Difference at NPL	Difference at RSMAS
$10°C$	-19 mK	below dew point
$20°C$	-16 mK	12 mK
$30°C$	-7 mK	-17 mK

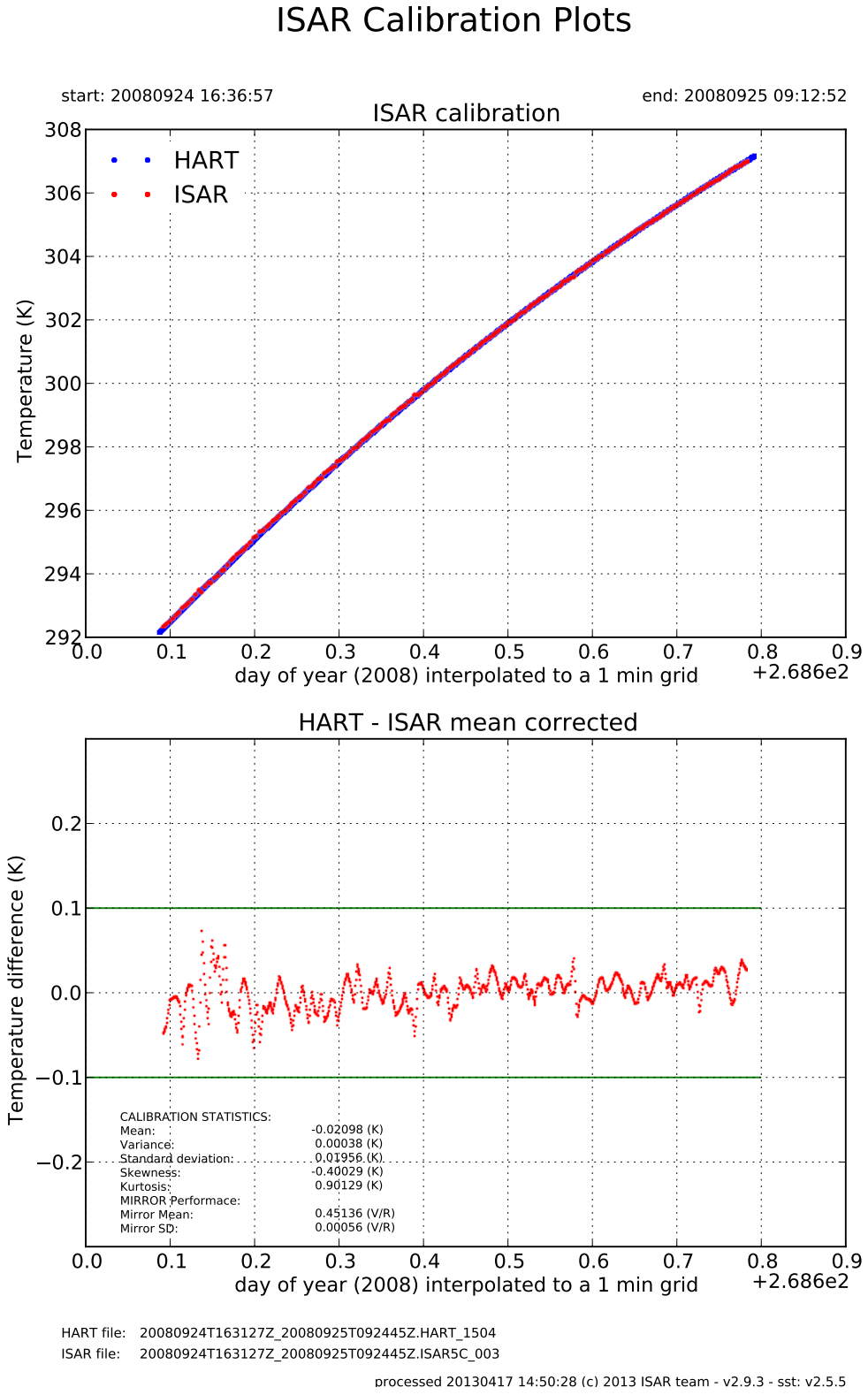


FIGURE 3.16: Calibration plot of ISAR 03 before Deployment 22

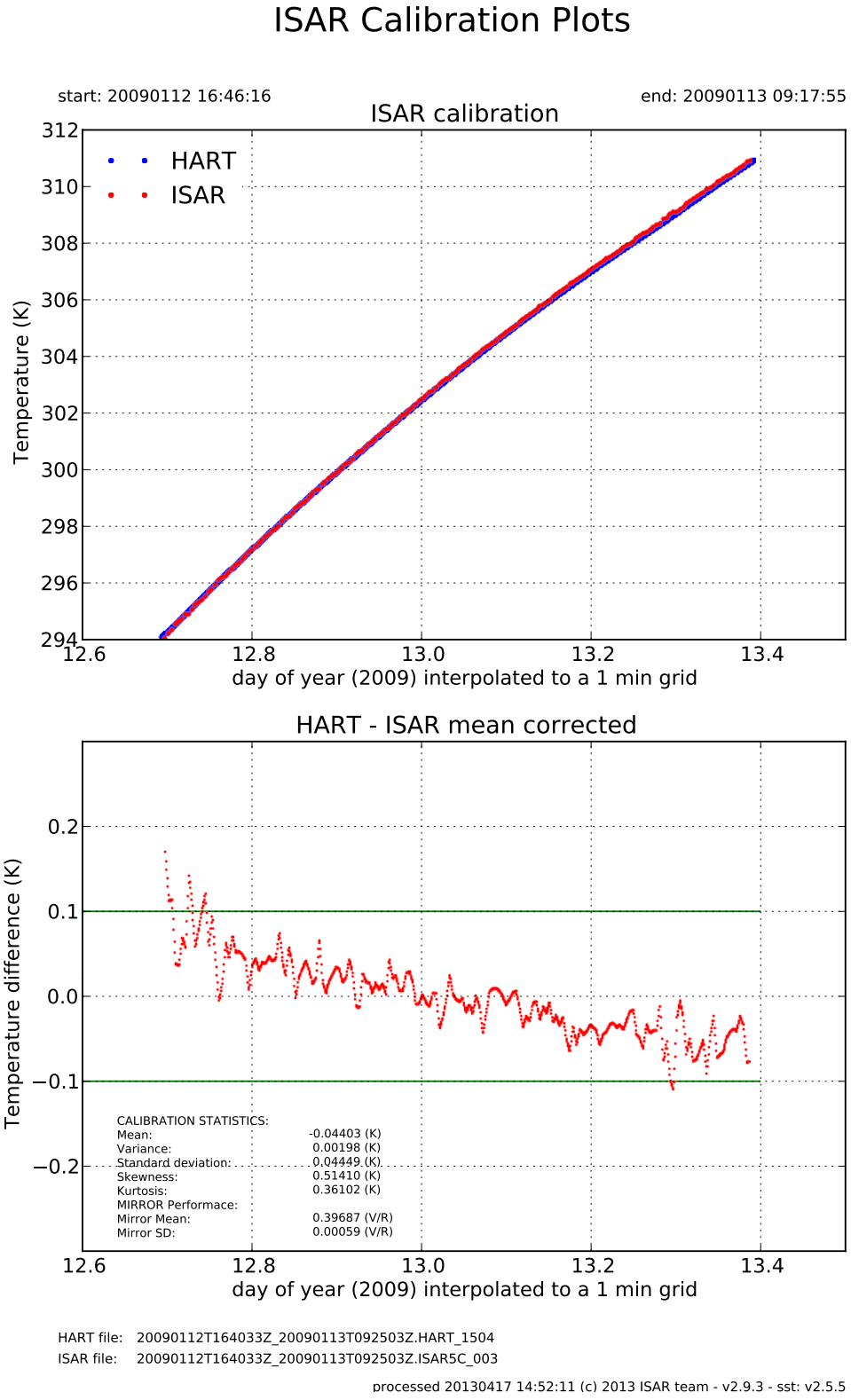


FIGURE 3.17: Calibration plot of ISAR 03 after Deployment 22

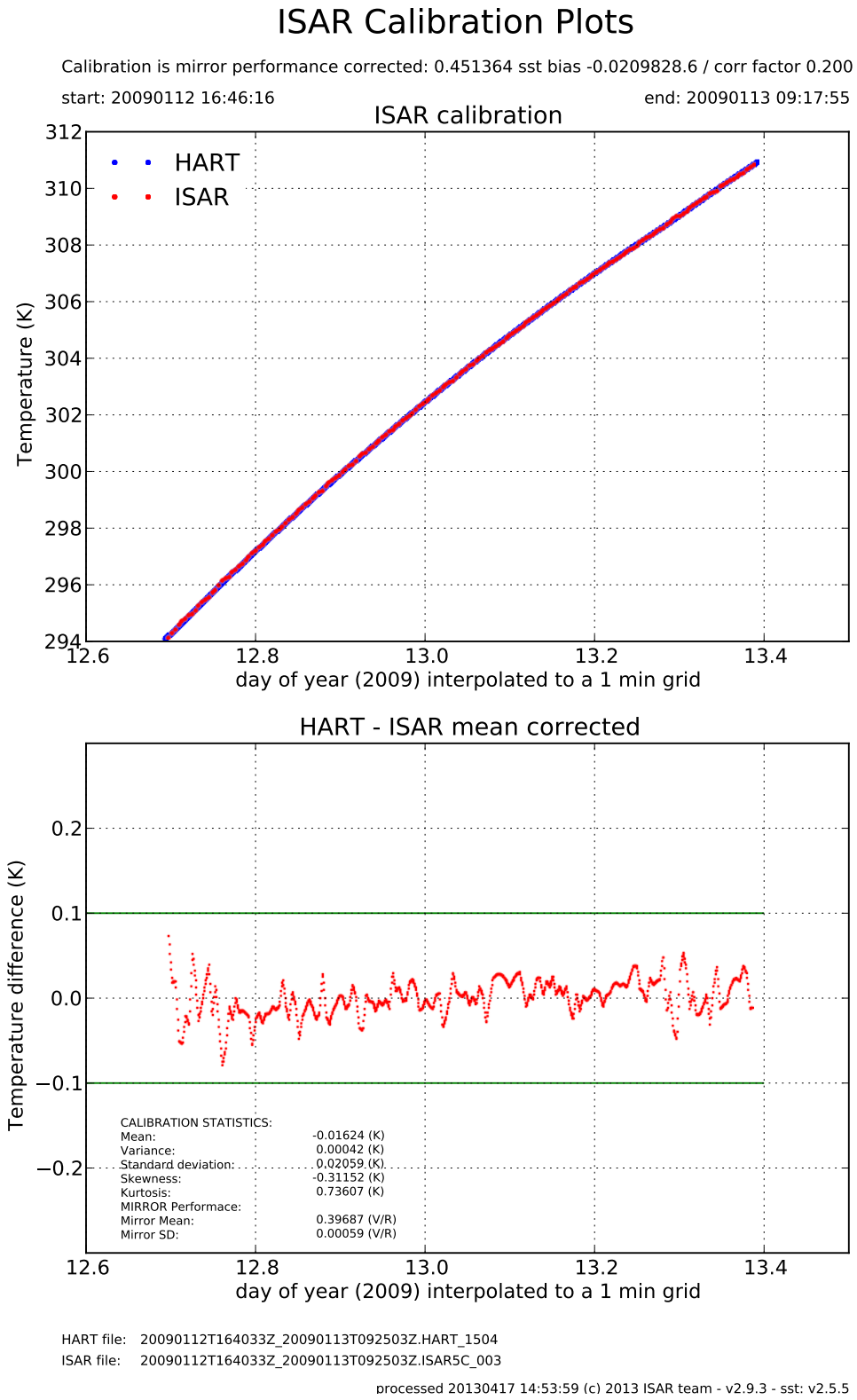


FIGURE 3.18: Calibration plot of ISAR 03 after Deployment 22 with the revised *SST* (see section 3.2.4) calculation applied

3.3 Standard ISAR configuration

The ISAR instrument can be configured by specifying adjustable parameters within a "config file" which resides on the on-board flash card. Most of the relevant instrument parameters such as scan drum angles, emissivities, radiance to temperature transformations, thermistor calibrations can be set in this config file. Table 3.5 shows a normal field configuration for the instrument's main parameters.

TABLE 3.5: ISAR standard field configuration

Parameter	Value	Unit	Comment
Scan angle 1	280	degrees	BB 1
Scan angle 2	325	degrees	BB 2
Scan angle 3	25	degrees	sky view
Scan angle 4	155	degrees	sea view
Scan angle 1 samples	30		BB 1
Scan angle 2 samples	30		BB 2
Scan angle 3 samples	10		sky view
Scan angle 4 samples	40		sea view
Emissivity	0.99164		sea water emissivity
BB emissivity	0.9993		internal black body emissivity
Rain gauge	0.064	V	MiniOrg 115 / 815
Rain gauge	1.2	V	Thies Clima

All the configuration information from the config file is stored as a header in the RAW ISAR data file. The header information is used by the ISAR *SST* processor to calculate all the relevant parameters for the *SST* estimation.

Unless otherwise stated, the ISAR SST_{skin} data that are used in subsequent chapters of this thesis were retrieved using the configuration defined in Table 3.5 .

3.4 Installation on the Pride of Bilbao

The *in situ* data used in this study were collected on the P & O ferry PoB which operated between Portsmouth in the UK and Bilbao on northern Spain until October 2010. The ISAR instrument was mounted on the starboard bridge roof as shown in Figure 3.19 and Figure **??**. Table 3.6 lists the instruments mounted on the PoB.

TABLE 3.6: List of instruments and their parameters on the PoB

Instrument	Manufacter	Height	Parameter	Unit	Uncertainty
ISAR	NOCS	35 m	SST_{skin}	K	0.1
ORG115	Optical Scientific Inc.	35 m	Precipitation	mmh^{-1}	6%
ORG815	Optical Scientific Inc.	35 m	Precipitation	mmh^{-1}	6%
Precipitation sensor	Thies Clima	35 m	Precipitation	mmh^{-1}	6%
CM 11	Kipp and Zonen	40 m	Short wave radiation	Wm^{-1}	1%
PIR	Eppley Laboratory	40 m	Long wave radiation	Wm^{-1}	5%
Windmaster Pro	Gill Instruments Ltd	43 m	Wind speed	ms^{-1}	1.5 %
			Wind direction	deg	2.0
SBE 48	Seabird	-5 m	SST_{depth}	K	0.002
RBR 1050	RBR Ltd.	-6 m	SST_{depth}	K	0.002
Hull Thermistor	YSI	-5 m	SST_{depth}	K	0.1

3.4.1 Sea Surface Temperature at depth

A number of SST_{depth} measurements were also made on the PoB together with the SST_{skin} measurements by the ISAR instrument. Those temperature measuring instruments whose data are used in later chapters of this thesis to identify the differences between the along track records of a shipborne radiometer and the type of temperature measurement conventionally delivered by ships of opportunity. They are briefly described in the following subsections. Furthermore the FerryBox suite of instruments (Hartman *et al.*, 2010) to sample the biochemical properties of the sea water, including its temperature at the engine intake, was installed on PoB. However because of quality control issues with the engine intake temperature as measured by the FerryBox instruments these were excluded from the subsequent analyses in chapter 5.

3.4.1.1 SBE 48

The Sea-Bird Electronics model 48 hull contact thermometer (SBE 48) (Seabird, 2010) is a through-the-hull thermometer installed at 5 m which measures the temperature of the sea water through the hull of the vessel and was specifically designed for use on

VOS. The SBE 48 temperature data have not exhibited any problems during the years 2006 - 2010 used for the analyses in chapter 5. The instrument was regularly calibrated, although because of the way the SBE 48 has to be installed (i.e.: inside the hull and in a warm room containing pumps and compressors about 70m from the bow) it does have an installation bias which was estimated to be $0.3 \pm 0.2\ K$ (see section 5.4).

3.4.1.2 Hull Thermistors

To investigate further the installation bias of the SBE 48 , and in an attempt to obtain information about the stratification of the water layer around the vessel, two sets of thermistors (YSI 46041, four thermistors each set) were installed on the port and starboard side of the PoB attached to the internal wall of the hull with a thermally conducting compound. The thermistors were placed at 2.5, 5 and 7.5 m nominal depth below the vessel's loading line and one thermistor for measuring the engine room compartment's air temperature was installed additionally. The port thermistor chain was situated approximately 15m further aft than the starboard thermistor chain, which is at the location of the SBE 48 instrument. The reason for mounting the port side thermistor chain further back was to measure the effects of the air temperature in the pump and compressor room on the measured SST_{depth}.

3.4.1.3 CPR temperature probe

The PoB deployed a Continuous Plankton Recorder (CPR) once a month for Sir Alister Hardy Foundation for Ocean Science (SAHFOS) to extend their plankton record (Reid *et al.*, 2003). To make maximum use of the CPR deployments a RBR 1050 temperature recorder (RBR, 2004) was installed with the help of SAHFOS on the CPR which provided a temperature reading at approximately 6 m water depth 60 m behind the ship in its wake. These data were used to estimate the installation bias of the SBE 48 as discussed in section 5.4.

3.4.2 Ancillary measurements

Together with the temperature measurements a number of atmospheric measurements were recorded on the PoB. These included wind speed and direction as measured by a Gill Windmaster 3 axis anemometer (Gill, 2010). The true wind speed was calculated by using the ISAR instrument's GPS information to derive the vessel's speed and direction and then 10m wind was calculated by using a static inversion profile and a air flow correction as described in Moat *et al.* (2006).

Short wave solar radiation was measured with a Kipp and Zonen CM11 (Kipp and Zonen, 2004). The longwave solar radiation was measured with a Eppley PIR (Fairall *et al.*, 1998) sensor and the thermopile voltage was corrected as described in Fairall *et al.* (1998). Furthermore air temperature and humidity were recorded with a Vaisala HMP243 (Vaisala, 1998).

The ancillary data from the PoB was not used explicitly for the work reported in this thesis but were available in case extra information was needed to explore anomalies of the temperature sensors.

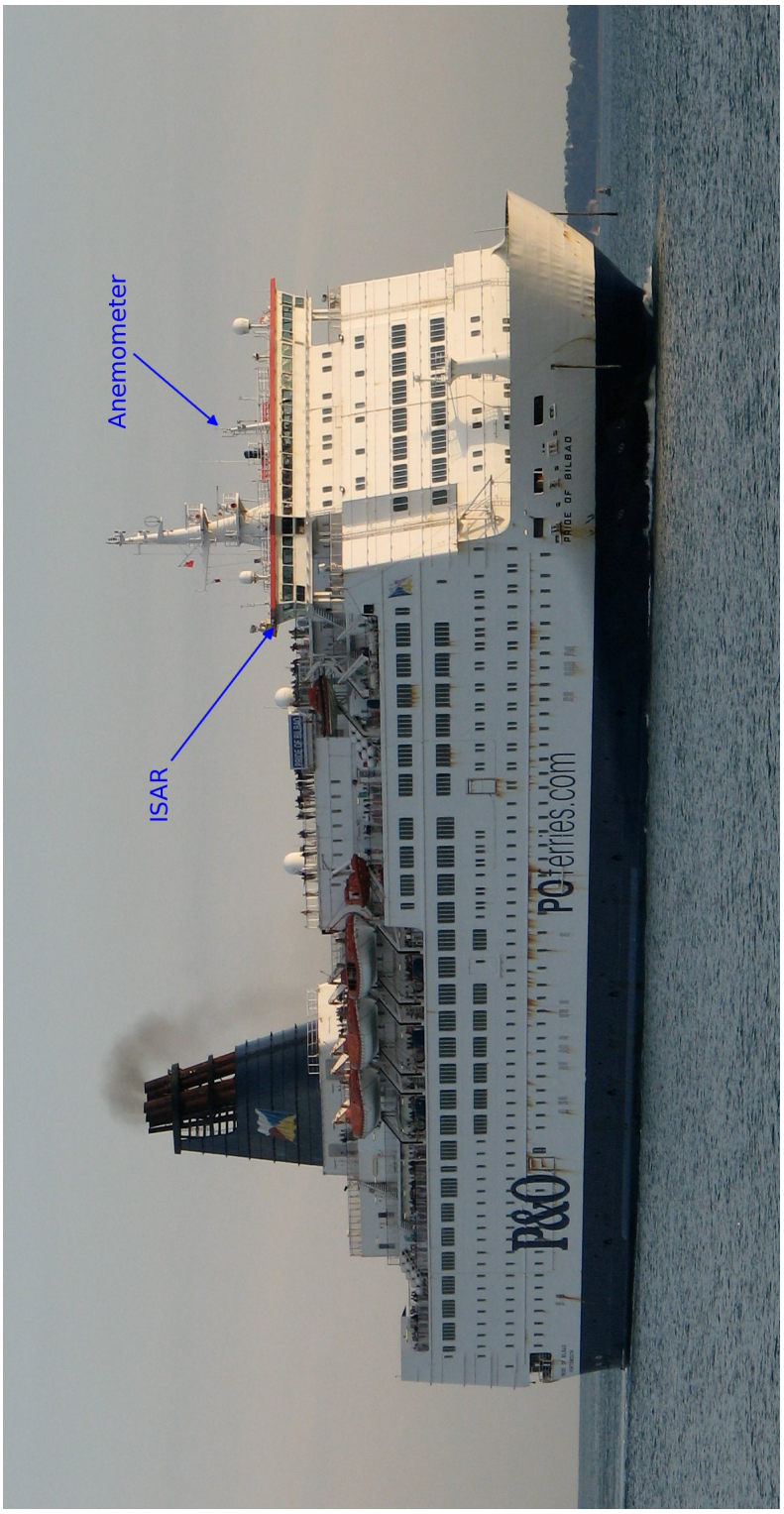


FIGURE 3.19: The instrument installation on the PoB

FIGURE 3.20: ISAR mounted on the starboard bridge roof on the PoB.

FIGURE 3.21: Close up of the ISAR mounted on the starboard bridge roof on the PoB.

3.5 Installation on the Cap Finistere

In October 2010 the ISAR instrument was moved to the Brittany Ferries operated CpF which covers a very similar route to that of the PoB, travelling between Portsmouth and Bilbao once a week and between Portsmouth and Santander twice a week. The higher frequency of the CpF compared to the PoB is due to the CpF being a newer ship and therefore capable of travelling at up to 28 knots, compared to the 20 knots maximum speed of the PoB.

The installation on CpF is a lot simpler and only consists of the ISAR instrument. This is mainly a consequence of spatial constraints in the engine room, with no room for a ferry box systems or other temperature sensors and due to a lack of accessible wiring to the bridge roof for the meteorological sensors.

Figures 3.22, 3.23 and 3.24 shows the ISAR installed on the starboard side bridge walkway on the CpF.

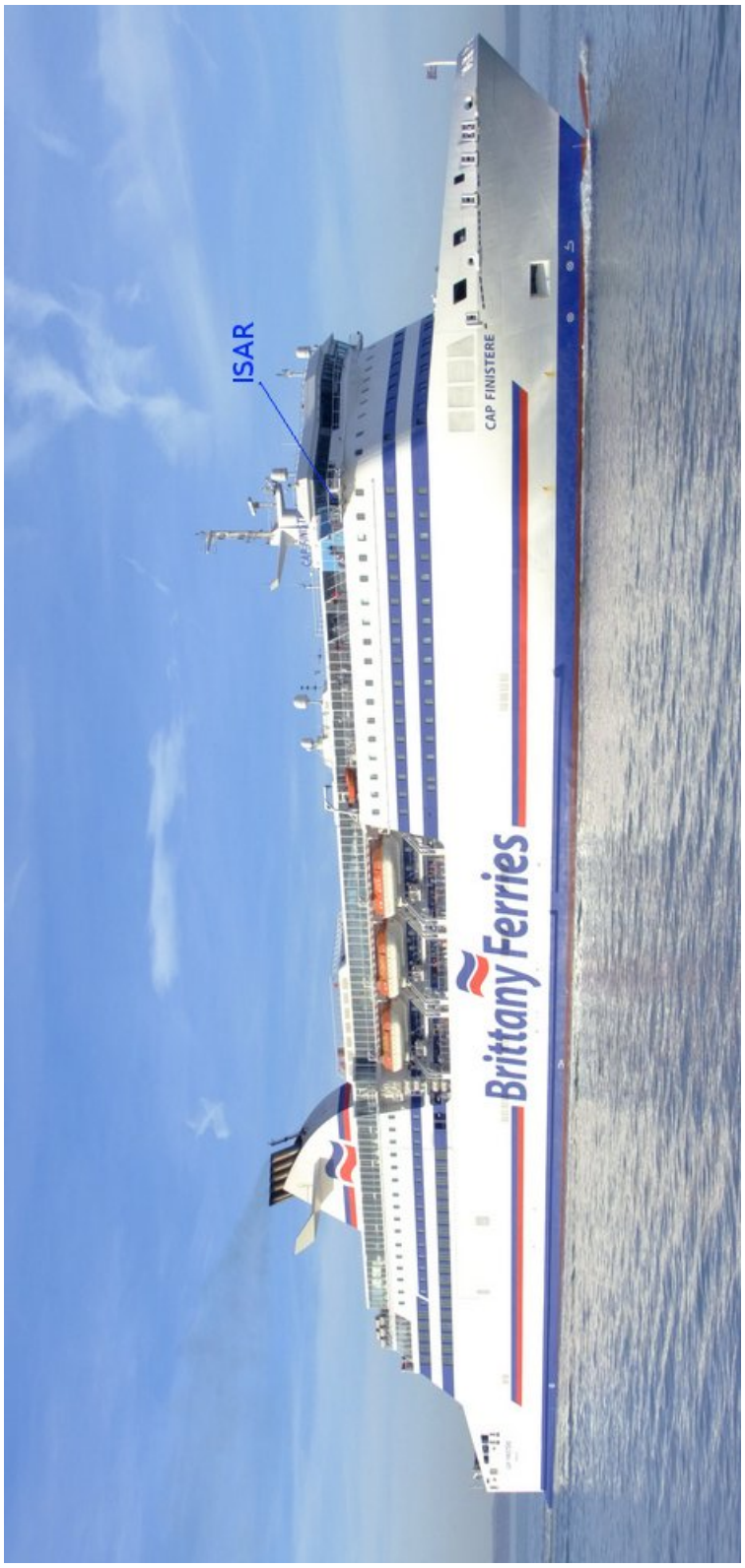


FIGURE 3.22: Cap Finistere with the ISAR location labelled.

FIGURE 3.23: ISAR mounted on the starboard bridge walkway on the CpF

FIGURE 3.24: Close up of the ISAR close up on the CpF

3.6 Overview of ISAR data

Figure 3.25 shows a hovmuller plot of the ISAR SST_{skin} data record from 2004 to 2011. Small gaps in the data are rain and sea spray events, big data gaps are when the ship was not running because of bad weather or the ship being at its annual re-fit. Deployments D5 and D24 are missing from the record because of instrument issues, causing them to fail the quality assurance procedure as described in section 3.2.5.2.

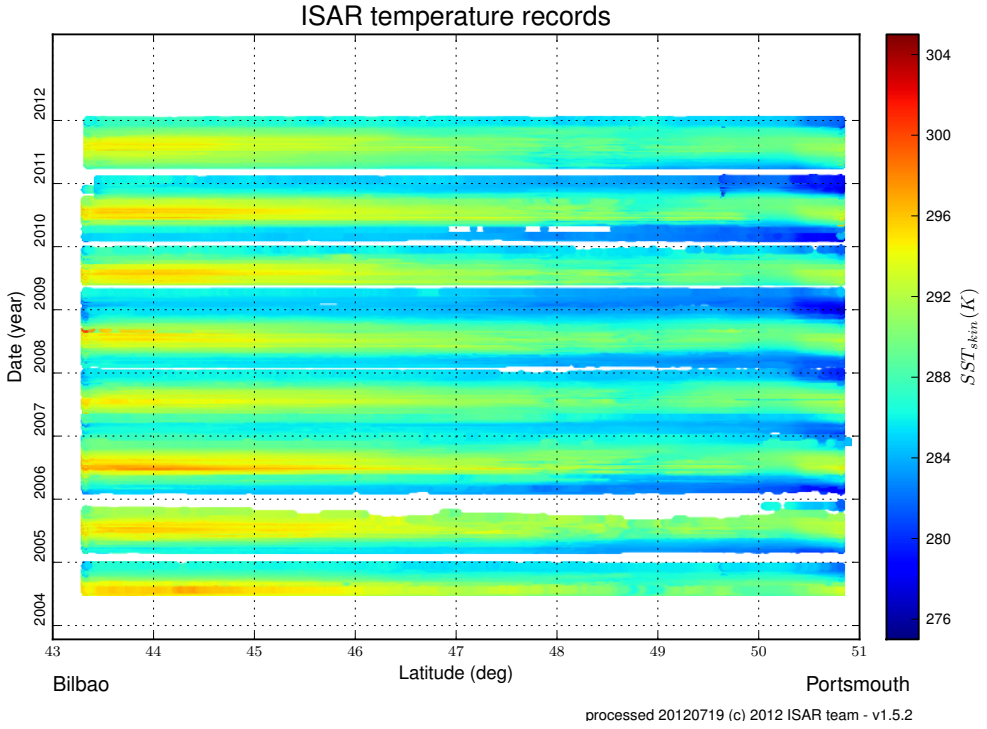


FIGURE 3.25: ISAR data collected from 2004 to 2011

Table 3.7 lists the ISAR deployments in the English Channel and Bay of Biscay.

A total of 437118 SST_{skin} measurements on 2546 deployment days have been acquired between March 2004 and July 2012. On average this equates to 171.4 SST_{skin} measurements per day which is equivalent to one SST_{skin} measurement every 8.4 minutes. This average measurement value includes the times ISAR was not measuring the ocean because of rain or sea spray. In essence ISAR is producing slightly more data than a buoy would if the data would be recorded every 10 min. However a buoy would produce

a more regular data record, as ISAR does close its shutter and does not measure the surface in poor weather conditions.

The data in table 3.7 shows some of the improvements made to the ISARs with the experience gained through the deployments. For example the quality assurance (QA) exclusion of some of the data of deployment 8 resulted in the development of the revised radiometric calculation as described in section 3.2.4. Another example is the relatively poor amount of data at the end of 2005 (see figure 3.25) which was due to a failing rain gauge which was replaced for deployment 9 and subsequently the rain gauge type was changed to a more robust model. The two shutter failure experiences (deployment 21 and 34) resulted in a change of the maintenance procedure (deployment 21) and a design change for the shutter sprocket (deployment 34) to avoid similar problems in the future.

Overall the ISAR has been reliable instrument for measuring SST_{skin} which is demonstrated by the number of measurements acquired over 8 years which makes the ISAR SST_{skin} record one of the longest *in situ* SST_{skin} records in the world.

At present (Spring 2012) there are six ISARs in existence that are owned by a number of institutions. ISAR 001 and 004 are owned by the University of Miami, ISAR 002 and 003 are owned by the University of Southampton and were used for this study, ISAR 005 is owned by the Ocean University in Qingdao and ISAR 006 is owned by Japan Aerospace Exploitation Agency (JAXA).

TABLE 3.7: List of ISAR deployments from 2004 to 2012 in the English Channel and the Bay of Biscay.

No	Ship	ISAR	Start Date	End Date	Days	No of *SST*	*SST* / Day	Days open	Comment
1	PoB	2	17.03.04	10.04.04	24	2374	99.0	8.4	
2	PoB	2	27.04.04	15.06.04	49	7658	156.3	25.6	Design change
3	PoB	3	16.07.04	16.10.04	92	15201	165.2	38.9	
4	PoB	2	16.10.04	21.12.04	66	11746	178.0	28.1	
5	PoB	3	01.02.05	15.03.05	42	-	-	0.8	QA *SST* rejection
6	PoB	2	15.03.05	22.06.05	99	15062	152.1	38.3	
7	PoB	3	22.06.05	20.09.05	90	10520	116.9	29.9	
8	PoB	2	20.09.05	13.12.05	84	1786	21.3	5.4	QA excluded data[2]
9	PoB	3	13.12.05	06.01.06	24	4267	177.8	14.6	
10	PoB	2	21.02.06	10.05.06	78	9411	120.7	22.2	
11	PoB	3	10.05.06	09.08.06	91	14513	159.5	51.9	
12	PoB	2	09.08.06	06.11.06	89	18544	208.4	56.8	
13	PoB	3	06.11.06	04.01.07	59	10667	180.8	35.4	
14	PoB	2	08.02.07	08.05.07	89	14124	158.7	47.6	
15	PoB	3	08.05.07	04.06.07	27	97	3.6	0.3	Electronics failure
16	PoB	2	04.06.07	07.08.07	64	16671	260.5	47.7	
17	PoB	3	07.08.07	25.11.07	110	16017	145.6	53.8	
18	PoB	3	29.11.07	07.01.08	39	4629	118.7	15.6	
19	PoB	3	21.02.08	28.05.08	97	11556	119.1	38.4	
20	PoB	3	31.05.08	01.09.08	93	19523	209.9	65.6	
21	PoB	3	04.09.08	16.09.08	12	2264	188.7	7.4	Shutter failure
22	PoB	3	01.10.08	06.01.09	97	16566	170.8	55.1	
23	PoB	3	04.02.09	15.04.09	70	11767	168.1	37.0	
24	PoB	2	15.04.09	10.06.09	56	-			QA *SST* rejection[3]
25	PoB	3	10.06.09	24.08.09	75	19942	265.9	67.6	
26	PoB	2	24.08.09	27.08.09	3	-			QA *SST* rejection[4]
27	PoB	3	27.08.09	16.12.09	111	12522	112.8	40.2	
28	PoB	3	19.02.10	21.05.10	91	4471	49.1	14.5	
29	PoB	2	21.05.10	28.05.10	7	-			QA *SST* rejection[5]
30	PoB	3	28.05.10	29.09.10	124	30443	245.5	104.1	
31	CpF	3	26.10.10	25.01.11	91	21454	235.8	44.2	
32	CpF	3	15.04.11	20.07.11	96	26178	272.7	67.1	
33	CpF	3	22.07.11	28.10.11	100	32214	322.1	82.9	
34	CpF	3	02.11.11	01.12.11	29	6534	225.3	16.5	Shutter failure
35	CpF	3	08.12.11	26.01.12	49	11108	226.7	27.9	
36	CpF	3	28.02.12	13.06.12	106	29723	280.4	76.6	
37	CpF	3	20.06.12	13.07.12	23	7566	329.0	19.1	End CpF

[2]Data after day 308 was rejected
[3]New electronics trial
[4]New electronics software update trial
[5]New electronics software update trial

3.7 AATSR - ISAR validation data

As mentioned in section 1.1 the context of this thesis is the validation of AATSR using ISAR which the author has been responsible for. While the validation work is not part of the thesis it provides the starting point for chapter 6 and the paper (Wimmer *et al.*, 2012) describing the work is included in appendix D. Wimmer *et al.* (2012) is one of the first papers showing the benefits of using *in situ* radiometric measurements along a constantly repeated route. A short excerpt of the results for AATSR dual view data from Wimmer *et al.* (2012) is shown in table 3.8. The "grade of coincidence" is specified in table 6.1 and is the temporal and spatial window in which data AATSR and ISAR are considered to be coinciding.

TABLE 3.8: Statistics for $match - up$ pairs (AATSR CV 5 filtered dual-view *SST* - ISAR) from 7th December 2005 to April 2009, showing the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the range of sea temperatures spanned by the match-up database. From Wimmer *et al.* (2012)

Grade of coincidence	Mean bias, AATSR - ISAR	Standard deviation	No. of Matches	No. of Overpasses	Min temp, $^{\circ}C$	Max temp, $^{\circ}C$
2 channel view (Dual-2)						
1	-0.03	0.30	136	36	6.4	18.4
2a	-0.02	0.60	209	57	6.4	20.4
2b	-0.04	0.39	549	58	6.0	21.5
3	-0.05	0.54	784	68	5.1	22.0
4	0.02	0.76	2372	131	4.8	24.1
3 channel view (Dual-3)						
1	0.00	0.23	326	48	5.9	22.1
2a	0.00	0.32	411	68	5.9	22.1
2b	-0.01	0.26	1137	62	5.9	22.6
3	-0.03	0.32	1423	87	5.9	22.6
4	-0.05	0.38	4299	140	4.1	24.9

Wimmer *et al.* (2012) did not only show that AATSR was performing well within its design parameters it also produced much higher $match - up$ numbers than previous publications. For example Corlett *et al.* (2006) only found 30 $match - up$ pairs between

AATSR and *in situ* radiometers from 2002 to 2003 for a grade 2b $match - up$ window. However the fairly simplistic approach of using fixed $match - up$ window does have a number of limitations which led to the development of the $match - up$ indicators as described in chapter 6.

Chapter 4

ISAR uncertainties

4.1 Introduction

In order to maximise the useful role of the ISAR measurements in validating satellite SST data, each measurement should come with an uncertainty where the uncertainty expresses the confidence we have that the measurement is correct within certain margins (Bell, 2001). The analysis of uncertainty is an estimation of the doubt one has in the measurement whereas an error analysis would mean we already know the true value and can therefore correct our measurement to the true value. In the rest of this chapter the estimated uncertainty for a variable implies that the estimate of the variable differs from its true value by less than the stated uncertainty in 95% of cases. This chapter will first analyse the individual contribution to uncertainty from each of the components of ISAR and then assess how these interact to determine an overall uncertainty for each SST_{skin} measurement as estimated by ISAR.

There are two general ways of characterising uncertainties:

- Type A: Uncertainties which must be estimated by using statistics, sometimes also referred to as random, since the uncertainties can be reduced by increasing the number of samples used in producing a single data record.

- Type B: Uncertainties estimated from knowledge of component behaviour or other information, sometimes also referred to as systematic, because they are not reduced by obtaining more samples.

A further way of characterising the uncertainties is to split them into:

- Measurement uncertainty: The uncertainty associated with the typical variability of the measured property. e.g.: For ISAR the variability of the Brightness Temperature of the sea view (BT_{sea}) and Brightness Temperature of the sky view (BT_{sky}).
- Instrument uncertainty: The uncertainty the measuring instrument introduces regardless of the measured property.

4.2 An uncertainty budget

To estimate an uncertainty for each SST_{skin} measurement we first have to analyse all the contributing elements of the ISAR instrument. Figure 4.1 shows a general outline of the ISAR SST processor and how the individual instrument components contribute to the estimation of the total SST uncertainty. Table 4.1 shows each instrument element, its uncertainty estimate, the uncertainty type and for Type B uncertainties the section reference explaining the estimation of that uncertainty.

To estimate an uncertainty budget from the individual contributions as listed in table 4.1 the convectional approach would normally be to add all the uncertainties together.

$$\sigma^2 = \sum \sigma_i^2 \tag{4.1}$$

This was not not followed. Such an approach only works where uncertainties are propagated in a linear fashion. The ISAR instrument however is a self calibrating radiometer where some of the uncertainties become reduced by the internal calibration process. Therefore, to estimate uncertainties for the ISAR instrument we attached the uncertainty to each of the contributing items as shown in table 4.1 and propagated these

TABLE 4.1: List of Uncertainties for ISAR

Item	Uncertainty	Unit	Type	Reference
Detector linearity	$< 0.01\%$	K/month	B	4.2.3.5
Detector noise	~ 0.002	Volts	A	N/A
Detector accuracy	± 0.5	K	B	4.2.3.5
ADC	± 1	LSB	B	4.2.3.3
ADC accuracy	$\pm 0.1\%$	Range	B	4.2.3.3
ADC zero drift	± 6	$\mu V\ /^{\circ}C$	B	4.2.3.3
Reference voltage 16 bit ADC	± 15	mV	B	4.2.3.3
Reference voltage 12 bit ADC	± 20	mV	B	4.2.3.3
Reference resistor	1	%	B	4.2.3.3
Reference resistor temperature coefficient	± 100	ppm/$^{\circ}C$	B	4.2.3.3
Black Body emissivity	± 0.000178	Emissivity	B	4.2.3.2
Sea surface emissivity	± 0.07	Emissivity	B	4.2.3.1
Steinhart-Hart approximation	± 0.01	K	B	4.2.3.4
Radiate transfer approximation	± 0.001	K	B	4.2.3.5
Thermistor	± 0.05	K	B	4.2.3.3
Thermistor noise	~ 0.002	Volts	A	N/A

uncertainties through the ISAR *SST* processor to obtain the total instrument uncertainty.

To propagate the uncertainties through the ISAR *SST* processor (figure 4.1) a linear error propagation model based on the uncertainty package in python (Lebigot, 2012) was used. Section 4.2.1 gives a brief description of the python uncertainty package used.

4.2.1 Uncertainty propagation

Lebigot (2012) states that mathematically, numbers with uncertainties are, in this package, probability distributions. They are not restricted to normal (Gaussian) distributions and can be any kind of distribution. These probability distributions are reduced to two numbers: a nominal value and a standard deviation. Furthermore the uncertainty package calculates the standard deviation of mathematical expressions through the linear approximation of error propagation theory. This is why this package also calculates partial derivatives. The standard deviations and nominal values calculated by the package are thus meaningful approximations as long as the functions involved have precise linear expansions in the region where the probability density of their variables is the largest. It is therefore important that uncertainties be small. Mathematically, this means that the

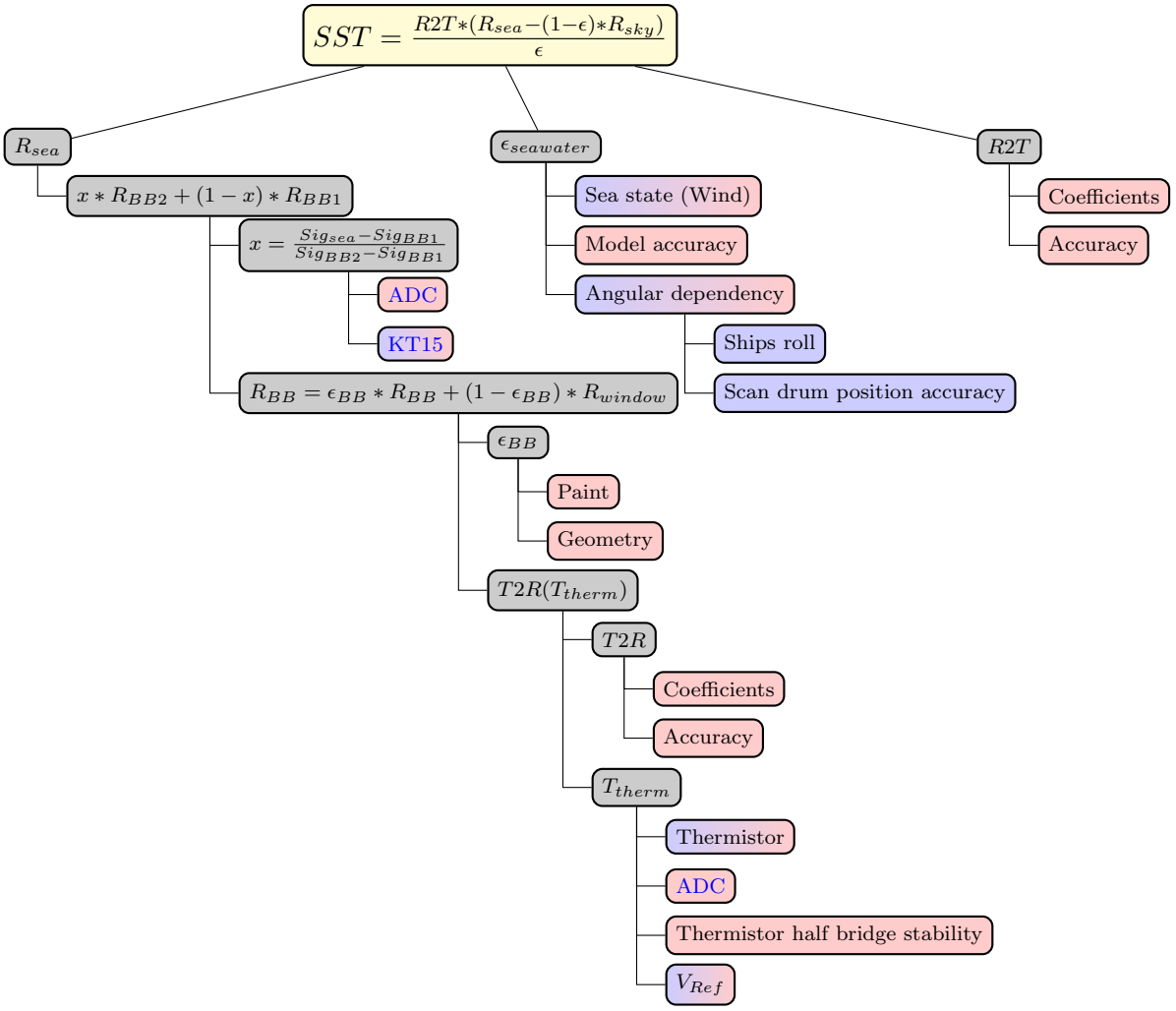


FIGURE 4.1: Flow chart of the ISAR *SST* processor. R_{sky} follows the same path as R_{sea}. Boxes coloured in blue show Type A uncertainties, boxes coloured in red show Type B uncertainties and boxes in red and blue show that the particular box has both Type A and Type B uncertainties.

linear terms of functions around the nominal values of their variables should be much larger than the remaining higher-order terms over the region of significant probability.

The python uncertainty package was applied using the mode and the variance of each parameter relevant to the ISAR *SST* calculation. This was done in order to reach the most statistically robust solution for the linear error model.

4.2.2 Type A uncertainties

The two main Type A uncertainties for the ISAR instrument are contributions by the KT15 detector on the one hand and by the thermistors on the other hand. The uncertainty estimation of the KT15 is fairly straight forward for the Type A component considered in this section, while the Type B component is discussed in section 4.2.3.5. In contrast the thermistors have several different individual components contributing to the overall uncertainty. However in this section we will only look at the Type A component. In both cases the Type A uncertainty was estimated by first calculating the variance of voltages as measured by the ADC over each view in one scan cycle. As a nominal value for the uncertainty propagation the mode for each view was used.

4.2.3 Type B uncertainties

4.2.3.1 Sea water emissivity

The emissivity of the sea surface does not only have Type B uncertainties, but also a Type A component, determined by the changing environmental conditions, such as the ship's roll or sea state. However, because the emissivity value itself is not determined by an ISAR measurement it seems more appropriate to discuss the sea water emissivity in the Type B uncertainty section.

Various studies have looked into the variability of sea water emissivity and its dependence on view angle and sea state. Wu and Smith (1997) and Masuda *et al.* (1988) present a theoretical approach for estimating the seawater emissivity, with Niclos *et al.* (2009) giving a simplified equation for the the atmospheric windows in the region. The emissivity model used by the ARC project (Embury *et al.*, 2012b) is based on Watts *et al.* (1996), Masuda *et al.* (1988) and Wu and Smith (1997), with some improvements from Newman *et al.* (2005). Newman *et al.* (2005) revisits the salinity and temperature dependence of ε, and while the salinity dependence is captured very well in Watts *et al.* (1996), Masuda *et al.* (1988) and Wu and Smith (1997), the temperature dependence is less well characterised. However this is mainly an issue in the $750 - 850 cm^{-1}$ wave

number region which is below the ISAR detectors filter region ($870 - 1050 cm^{-1}$) and at high viewing angles ($> 50°$).

Hanafin and Minnett (2005) and Niclos *et al.* (2005) estimated the surface emissivity from *in situ* measurements and Hanafin and Minnett (2005) found that at a view angle of 55° an error of up to 0.7 K can be introduced into the shipboard radiometer *SST* calculation. Both Hanafin and Minnett (2005) and Niclos *et al.* (2009) show that ε does not decrease as much with increasing wind speed as Watts *et al.* (1996), Masuda *et al.* (1988) and Wu and Smith (1997) predict. However Niclos *et al.* (2005) shows that the wind speed dependence of emissivity is near zero at wind speeds up to $10 ms^{-1}$. Masuda (2006) revisited the Masuda *et al.* (1988) and Wu and Smith (1997) calculations and added a SESR term which results in similar ε as Hanafin and Minnett (2005). The SESR term has very little effect below a 40° view angle and increases from a view angle of 50° showing a maximum value of 0.03 at an view angle of 80° to be added to ε.

To investigate the impact of the sea water emissivity on the overall uncertainty budget of the ISAR *SST* estimation a simple model of the ISAR *SST* estimation was used. The model in equation 4.2 is based on equation 3.18 and simply varies the input parameters in a range as outlined in table 4.2. The value ranges are typical ranges for the BT_{sky}, BT_{sea} and . The step value in table 4.2 was determined by the integration time of the model and has no effect on the result.

$$SST_{skin} = \frac{BT_{sea} - [1 - \varepsilon]BT_{sky}}{\varepsilon} \tag{4.2}$$

TABLE 4.2: List of parameter ranges used in the *SST* model

Parameter	Range	Step	Unit
BT_{sea}	270 - 320	0.5	K
BT_{sky}	190 - 290	0.5	K
ε	0.989 - 0.991	0.0001	

Figure 4.2 and 4.3 show the output of the simple emissivity model. The standard deviation (*std*) is calculated over the emissivity range as stated in table 4.2 for each BT_{sky} and BT_{sea}. Both figures show that the uncertainty converges to 0.0 the closer

BT_{sky} and BT_{sea} are to each other. This is to be expected from equation 4.2 and fits well with the theory. The plots in figures 4.2 and 4.3 also show that the uncertainty introduced by the modelled ε is in a similar range to that measured by Hanafin and Minnett (2005).

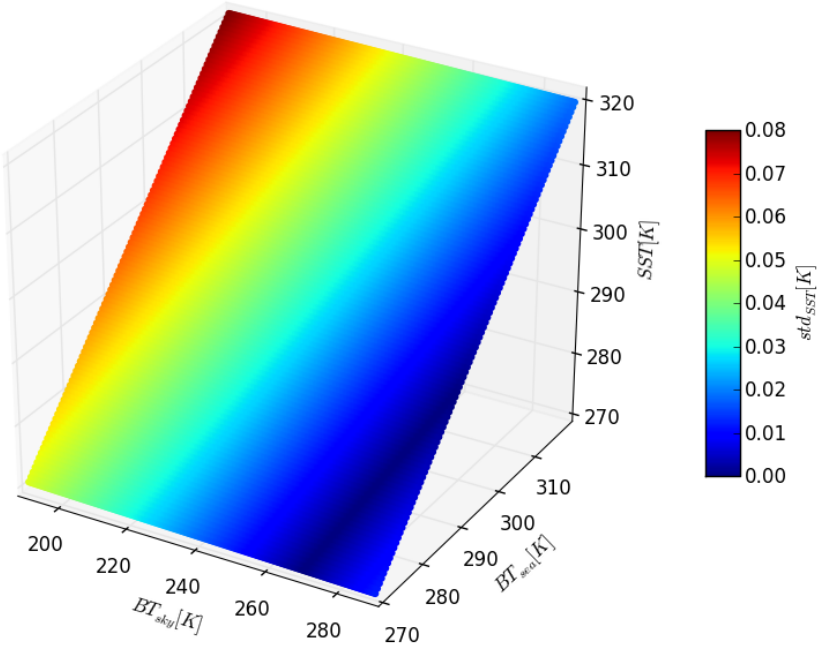


FIGURE 4.2: Plot of the effect the uncertainty in the seawater emissivity has on the SST processor uncertainty. Here BT_{sea} is the simulated Brightness Temperature of the sea view, BT_{sky} is the simulated Brightness Temperature of the sky view, SST the calculated Sea Surface Temperature and std_{SST} the resulting uncertainty of the SST estimation.

Because of the view angle dependence of ε the actual view angle has to be considered for the uncertainty budget. This was achieved by using ship roll angle as measured by the ISAR instrument's own pitch and roll sensor to calculate the actual view angle of the ISAR instrument. Donlon and Nightingale (2000) has also shown that because of the ship movement not only the changing emissivity but also the resulting mispointing of the sky view has to be considered in the uncertainty budget. Donlon and Nightingale (2000) give a deviation of 0.025 K for a sky view mispointed by 10 degrees. We used the Niclos *et al.* (2009) model to calculate the change in ε with changing view angles and calculated the Niclos *et al.* (2009) model over varying wind speeds from $0 - 20ms^{-1}$. The ε_0 values for Niclos *et al.* (2009) are calculated from Advanced Spaceborne Thermal Emission Reflection Radiometer (ASTER) (Baldridge *et al.*, 2009) seawater emissivity

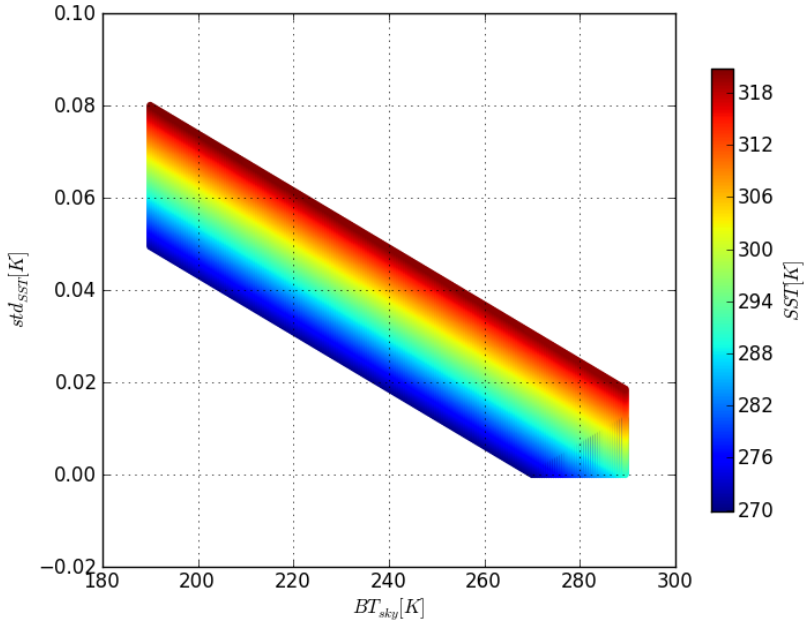


FIGURE 4.3: Plot of the change in uncertainty depending on BT_{sky} and SST. Where BT_{sky} is the simulated Brightness Temperature of the sky view, SST the calculated Sea Surface Temperature and std_{SST} the resulting uncertainty of the SST estimation.

values integrated over the ISAR spectral window. The ε_0 for the ISAR view angle of 25 degrees is calculated as 0.9916.

By taking all the above mentioned considerations into account the sea water emissivity could range from 0.92 to 0.9916. By looking at the ISAR data we get an uncertainty for the emissivity ranging from $\pm 10^{-4}$ to $\pm 10^{-7}$ depending on the conditions.

4.2.3.2 Internal black body emissivity

The internal ISAR BB cavities allow the instrument to calibrate the KT15 signal every scan cycle. Furthermore the thermistors in the internal BB are traceable to SI standards (NIST traceable calibration to $\pm 0.05K$, Measurement Specialties (2008)) and give the instrument its traceability. Figure 4.4 shows the geometry of the BB and the location of the three thermistors with which each black body is fitted. The BB use a re-entrant cone and a partially closed aperture design which, combined with a high emissivity surface finish (Nextel velvet black) and critical internal geometry, ensure that the black body cavities have an emissivity of > 0.999 in the thermal infra-red waveband (Berry, 1981).

Donlon *et al.* (2008) estimated the internal BB emissivity to be 0.9993. We revisited the calculation and estimated an uncertainty for the emissivity by taking ageing of the paint and the uncertainty of the Berry (1981) estimation into account. The result of this calculation gives an emissivity of 0.9993 ± 0.000178. This is lower than the analysis for the M-AERI which estimated the BB emissivity uncertainty as ± 0.0008 (Best *et al.*, 2003), however ISAR uses a re-entrant cone design and significantly closed aperture which is less susceptible to the ageing BB paint.

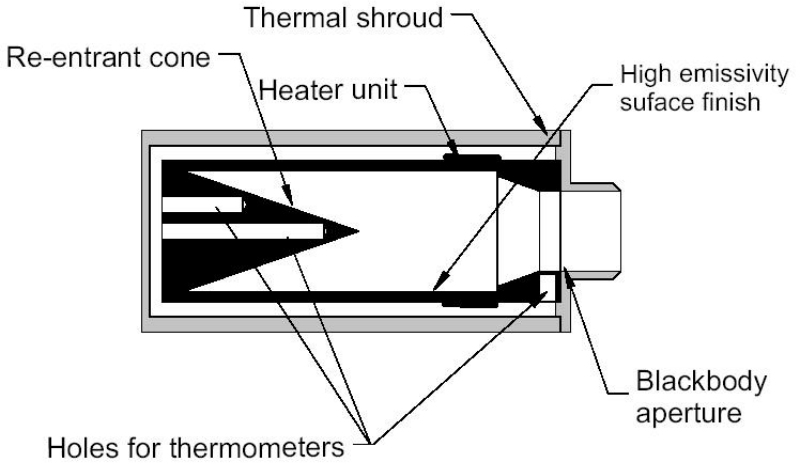


FIGURE 4.4: Section through the an ISAR calibration BB radiance cavity showing the re-entrant cone design, thermal shroud and location of thermistors used to determine the radiative temperature of the BB. The inner surfaces of the BB are coated with Nextel velvet black 811-21 paint. The emissivity of this design is 0.9993. Image from Donlon *et al.* (2008)

4.2.3.3 Thermistors

The thermistors used in the ISAR instrument are Yellow Springs Instrument Company (YSI) 46041 [1] The YSI 46041 is a super-stable (drift is $< 0.01K/100months$) interchangeable thermistor accurate to $\pm 0.05K$. The YSI 46041 is part of a half bridge which has a Bourns 4808 (Bourns, 2006) resistor network as the other part of the half bridge. The Bourns 4808 is a $10kOhm$ 1% resistor with a thermal coefficient of $\pm 100ppm\ /^{\circ}C$. The reference voltage to the thermistor half bridge is supplied by a Maxim MAX667 (Maxim, 2008) which has a line regulation uncertainty of $\pm 15mV$. The ADC is an Adam 4017 (Advantech, 1997) 8 channel 16 bit ADC with a software configurable range which

[1] YSI temperature division was purchased by Measurement Specialties from YSI Incorporated (Ohio, USA) in April 2006. However for the purpose of this thesis we still call the thermistor a YSI 46041.

was set to 5V. The accuracy is 0.1% of the range voltage (in the case of ISAR $5V$) and has a zero drift of $\pm 6\mu V\ /^{\circ}C$.

4.2.3.4 Resistance to Temperature approximation

A Steinhart-Hart approximation is used for the BB thermistors to estimate the measured temperature in the BB. The YSI 46401 data, as provided by Measurement Specialties (MS) (Measurement Specialties, 2008), was used to calculate a polynomial fit through the data as shown in figure 4.5 top panel. The middle panel in figure 4.5 shows the residuals for the Steinhart - Hart approximation in red and for the normal third order polynomial fit in blue. The bottom panel in figure 4.5 shows a tenth order fit for the estimation of the residuals which has difficulty replicating the noise in the residuals around $20^{\circ}C$. The noise around $20^{\circ}C$ is due to the coarse stepping of the original data supplied by Measurement Specialties (2008). The middle panel in figure 4.5 also shows that the Steinhart-Hart approximation has an increased uncertainty below $0^{\circ}C$ and above $75^{\circ}C$. While the upper limit is not very critical for this application the lower end has to be considered if ice surface measurements are made. For the purpose of this uncertainty analysis we used a fixed uncertainty for the Steinhart-Hart approximation of $\pm 0.01K$. While using a fixed uncertainty is not ideal, the bottom panel in figure 4.5 clearly shows that a polynomial fit through the residuals would not estimate the uncertainties around $20^{\circ}C$ correctly.

4.2.3.5 KT15

The detector used in the ISAR instrument is a Heitronics KT15.85D with a temperature range from $-100.0^{\circ}C$ to $50^{\circ}C$. The KT15 is linear in radiance but not in temperature, figure 4.6 shows the detector response in relation to the target temperature. The detector accuracy is ± 0.5 K + 0.7% of the temperature difference between detector case and the target temperature. The long term stability of the detector is better than $0.01\%K/month$ (Heitronics, 2000). The spectral response of a number of KT15 used for the existing ISAR instruments is shown in figure 4.7. Due to the differing spectral

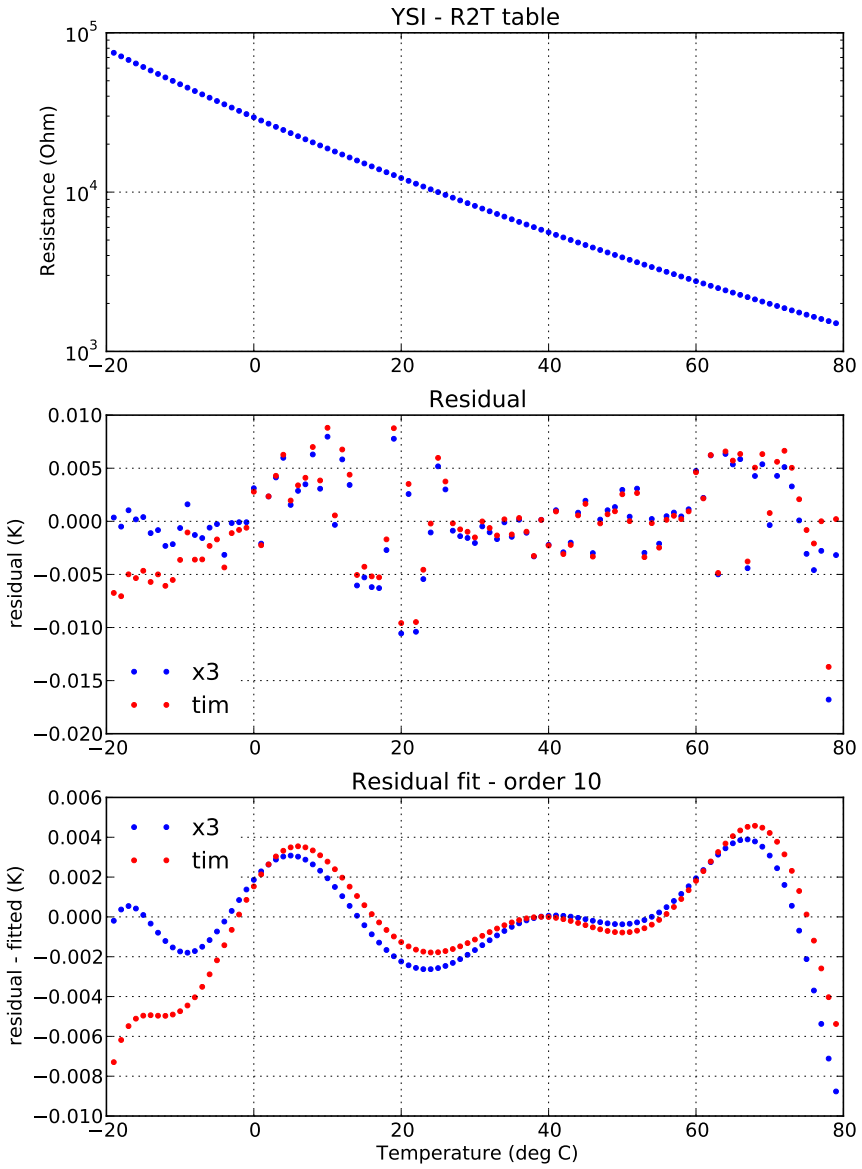


FIGURE 4.5: The YSI 46041 Thermistors resistance to temperature data is shown in the top panel; the middle panel shows the residuals of the residuals of the Steinhart-Hart approximation in red and a third order polynomial fit in blue and teh bottom panel shows a tenth order fit of the residuals shown in the middle panel.

responses of the KT15 the radiance to temperature conversion polynomials have been estimated separately for each KT15. The radiance to temperature conversion polynomials were estimated by Nightingale (2000) and their uncertainty is quoted as $\pm 1mK$.

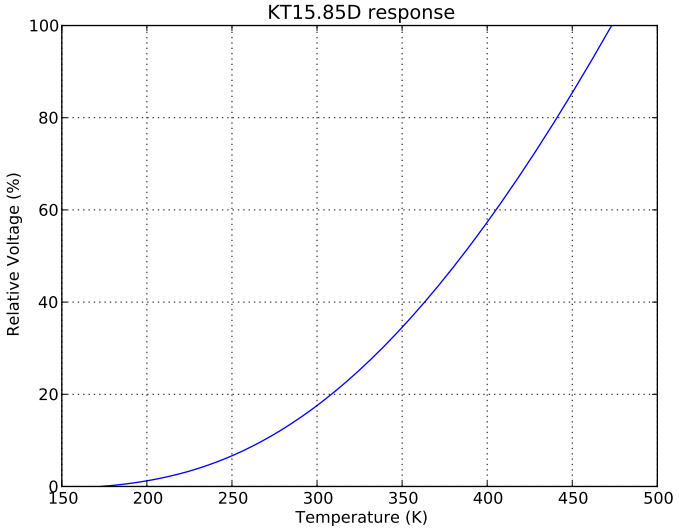


FIGURE 4.6: The KT15.85D detector response shown as relative voltage vs target temperature.

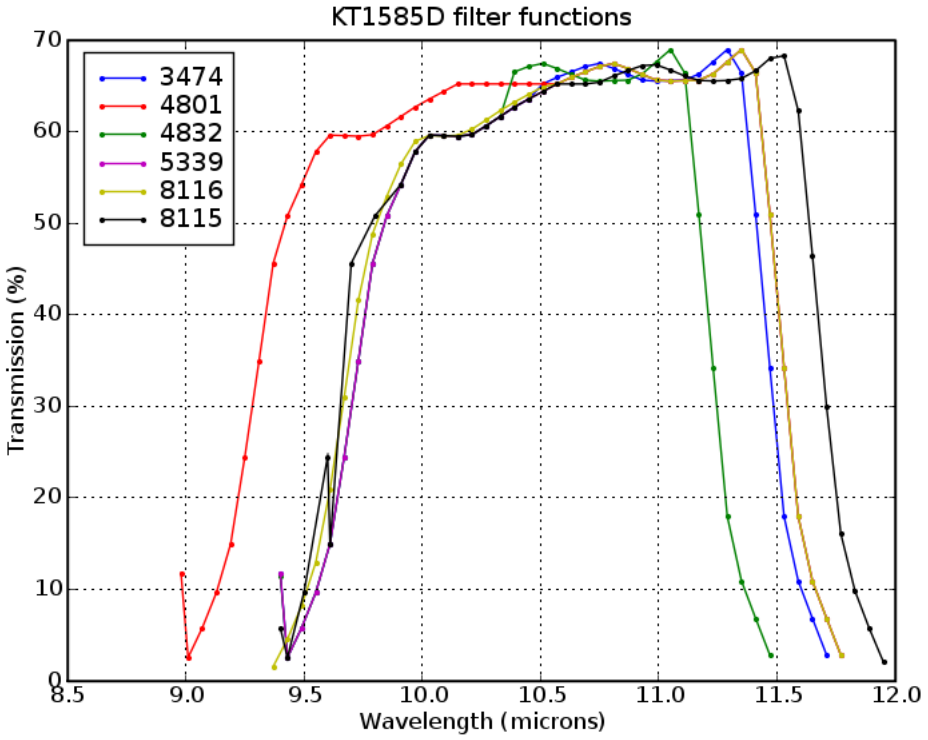


FIGURE 4.7: The KT15.85D detector filter response as used in ISAR. The numbers in the label box of the plot refer to the KT15.85D serial number as used in the different ISAR.

4.3 Estimating the ISAR uncertainty

After considering all the individual uncertainties contributing to the total ISAR instrument uncertainty the overall uncertainty is estimated by propagating the individual uncertainties through the ISAR *SST* processor as described in section 3.2.2 and 3.2.3 adding the uncertainties to each step:

(i) The thermistors uncertainties are calculated by first assigning the ADC uncertainty to the measured voltage. Then the resistance of the thermistor is calculated by taking the uncertainty of the reference voltage and the reference resistor into account. The resistance is converted into a temperature by using the Steinhart-Hart approximation taking the uncertainty of the approximation into account. The uncertainty resulting from this temperature calculation is then combined with the manufacturer quoted thermistor uncertainty of $\pm$ 0.05 K and the variability of the thermistor temperature during a BB target view to derive the thermistor uncertainty used in the temperature to radiance transformation.

(ii) For the detector signal the ADC uncertainty together with the KT15 temperature dependence, using the temperature of the KT15 case thermistor (which is calculated using the approach described above) and the variability over each target view is used to calculate the KT15 signal uncertainty.

(iii) Then the internal BB radiances with uncertainties are calculated by using the BB and window thermistor temperatures together with the internal BB emissivity uncertainty by using equation 3.7.

(iv) Now the sky and sea view radiance can be calculated by first calibrating the detector signal together with its uncertainties as described in equation 3.3 and 3.4 and then using the calibrated detector signal to calculate sky and sea radiance with the associated uncertainty from the BB radiances as shown in equation 3.6. The BB radiances for 3.6 are estimated as described in step one.

(v) ε is estimated by using the Niclos *et al.* (2009) model averaged over a wind speed range of 0 -20 $m\ s^{-1}$ with the view angle calculated from the maximum ship roll during

the sea and sky view.

(vi) The last step is calculating the *SST* with a total uncertainty by using the sea and sky radiances with uncertainty together with ε with uncertainty in equation 3.19.

Figure 4.8 shows the total uncertainty as calculated for ISAR data collected between 15.07.2011 and 22.07.2011 on the CpF. Also shown in figure 4.8 are the uncertainty split into Type A and Type B (bottom left panel) and into the instrument and measurement uncertainty (bottom right panel).

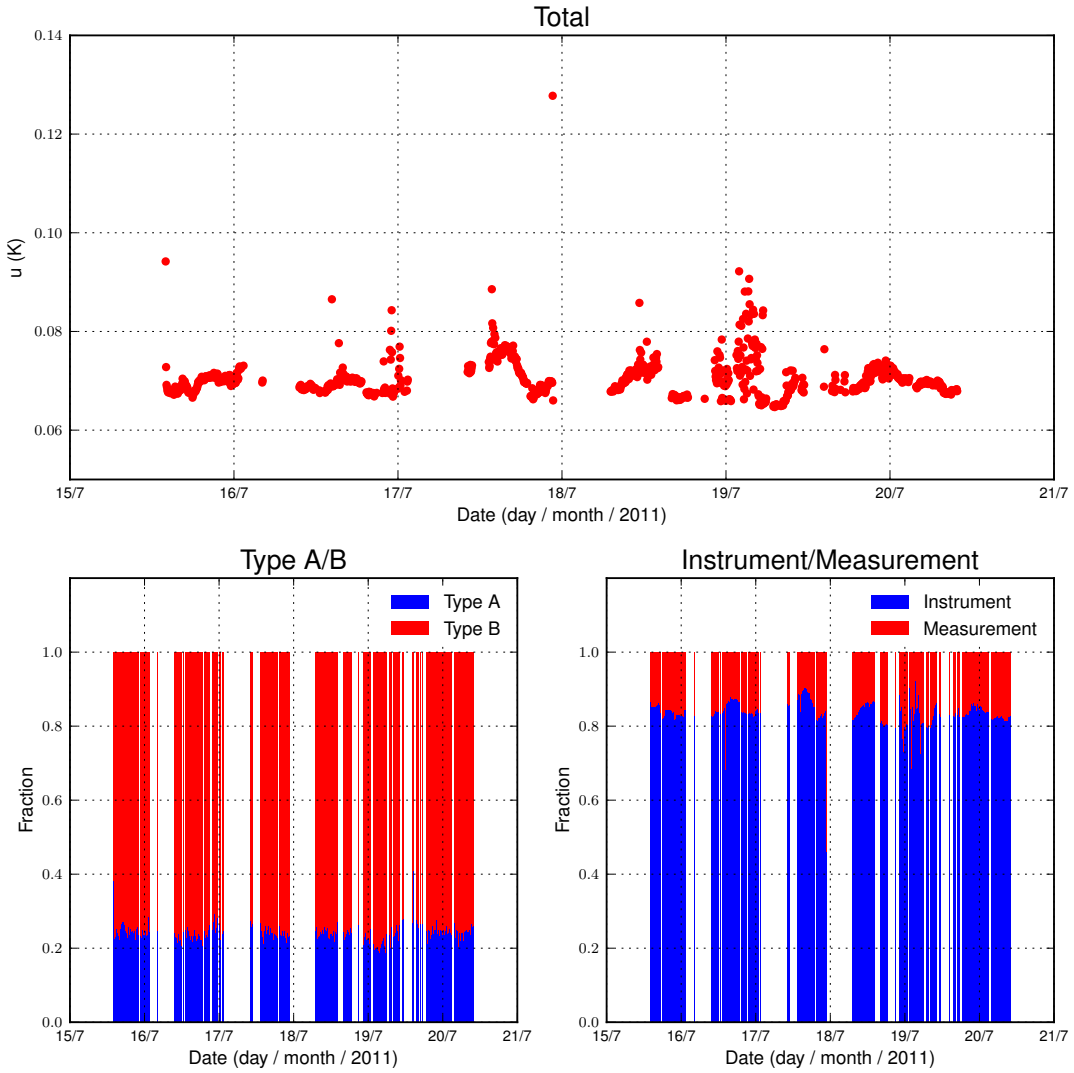


FIGURE 4.8: ISAR uncertainties for the data collected between 15.07.2011 and 20.07.2011 on the CpF. The top panel shows the measurement uncertainty. the bottom left panel shows the uncertainty split into fractions of Type A and Type B uncertainty and the bottom right panel shows the fraction of instrument and measurement uncertainty.

To estimate the Type A, Type B, measurement and instrument uncertainty the ISAR *SST* processor was run four times: one time each with one of the Type A, Type B or instrument uncertainty switched off and once with all parameters switched on for the total uncertainty. For the separation of Type A and Type B uncertainties, the definition as given in section 4.1, was followed. To separate the measurement and instrument uncertainty the target detector views (sea and sky view) uncertainty was set to 0.0 to obtain the instrument uncertainty. The measurement uncertainty was calculated as the difference between the instrument uncertainty and the total uncertainty.

The target uncertainty for ISAR was to be below 0.1 K as quoted in the instrument specification (Donlon *et al.*, 2008). While figure 4.8 shows that the total uncertainty is in most cases below 0.1 K, it is not always the case. Figure 4.9 shows the uncertainty data plotted against a few key parameter as well as a histogram of the uncertainty. The colouring of data points in blue and red in figure 4.9 is to distinguish between those cases where the uncertainty is associated with a data point where the BT_{sea} had a BT which lies between the ambient and hot BB (red dots) and those where the BT_{sea} had a BT lower than the ambient BB (blue dots). The colouring was applied in order to investigate whether there is a dependence of the uncertainty on what the measured sea temperature was in relation to the internal calibration target temperatures. And while figure 4.9 panel b shows a small dependency of the measured BT in relation to the ambient BB, it has no effect whether the measurement is between the two BB or not. The main driver for the small dependency shown in panel b is the emissivity of the internal BB. Figure 4.9 panel f shows that there is also a dependency of the uncertainty on the BT_{sky}, especially at very low BT_{sky}, which is expected as this shows the effect of the ε. Panel c in figure 4.9 shows that a lot of the higher uncertainty values occur in port as well as in the Bay of Biscay.

Figure 4.10 shows a geographical map of the same data as panel c in figure 4.9 however some of the high uncertainties in port are not so visible in figure 4.10 because the plotting of the dots on top of each other masks some of the higher uncertainties. The ISAR uncertainty does not seem to be dependant on the BT_{sea} or the SST_{skin}, as confirmed by panels e and h respectively in figure 4.9.

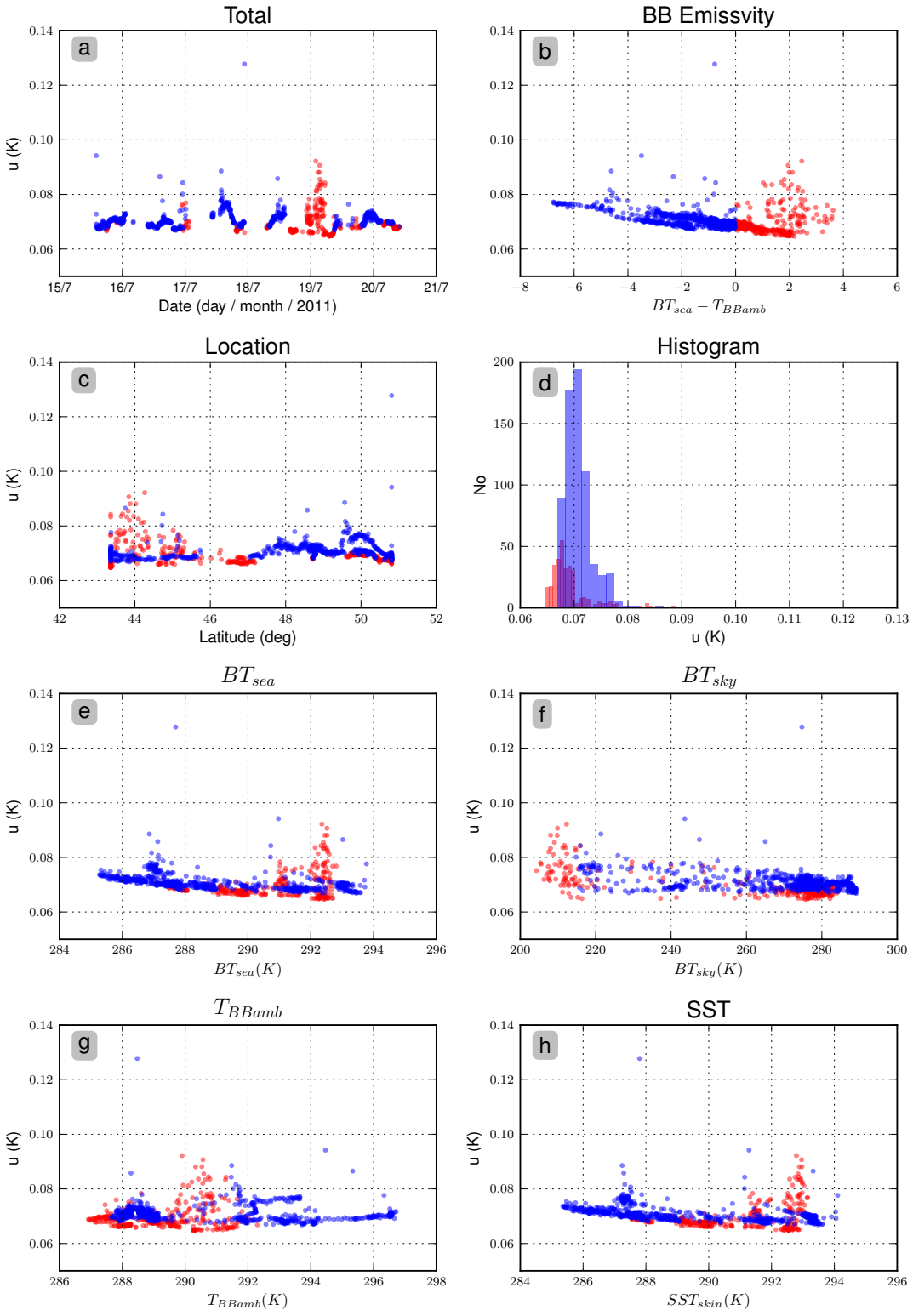


FIGURE 4.9: ISAR uncertainties for the data collected between 15.07.2011 and 20.07.2011 on the CpF. Panel a shows the total uncertainty, panel b the uncertainty in over the $BT_{sea} - T_{BBamb}$ difference, panel c shows the uncertainty plotted against the latitude and panel d a histogram of the uncertainty. Panel e shows the uncertainty plotted against the BT_{sea}, panel f shows the uncertainty plotted against the BT_{sky}, panel g shows the uncertainty plotted against the temperature of the ambient BB and panel h shows the uncertainty plotted against the SST_{skin}. The blue dots represent data where BT_{sea} is colder than the ambient BB and red dots show data where BT_{sea} is between the ambient BB and the hot BB.

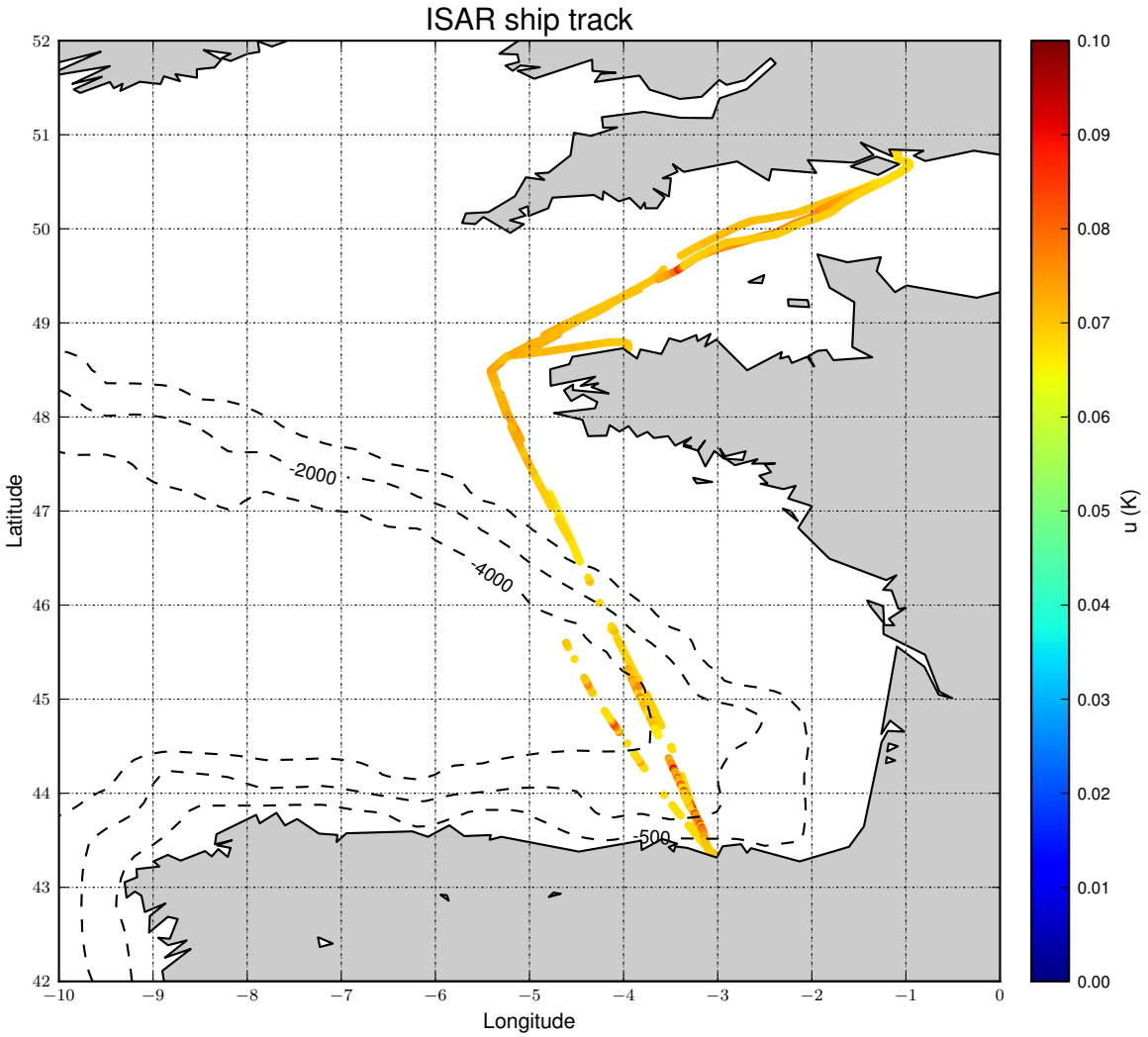


FIGURE 4.10: ISAR uncertainties for the data collected between 15.07.2011 and 20.07.2011 on the CpF plotted along the ship track.

4.4 Summary

In this section we have shown an approach to estimate an uncertainty for each SST_{skin} measurement cycle for the ISAR instrument. While static uncertainty budgets of varying detail did exist for *in situ* radiometers such as M-AERI (Best *et al.*, 2003) and ISAR (Donlon *et al.*, 2008) neither of those references consider how to attach uncertainty estimates to each SST_{skin} measurement. While a static uncertainty is sufficient to satisfy general measurement principles, Committee on Earth Observation Satellites (CEOS) and QA4EO have produced guidelines which ask for an uncertainty attached to each SST_{skin} measurement. This is especially important for satellite validation to ensure that the satellite uncertainty budget is not increased by wrongly attributing some of the poorly estimated *in situ* radiometer uncertainty to the satellite measurement.

The uncertainty budget for the ISAR as described here is the first attempt to do so and is in some aspects highly dependent on previous work on infra-red radiation uncertainties presented in the literature. This is especially evident in the case of the internal BB where independent thermistor calibration as well as measuring the BB emissivity together with witness samples of the black body paint ageing process would probably lead to a more robust estimation of the uncertainty. Also further work on the instrument mispointing uncertainty, as shown by Donlon and Nightingale (2000), would be useful. However because the ISAR data used for the uncertainty estimation are collected on a semi-operational basis for the AATSR validation contract[2] (Wimmer *et al.*, 2012) changes to instrument configuration, such as scanning over a number of view angles, could not easily be accommodated. The main reason for running the ISAR in a specific configuration, see section 3.3 was that this configuration is optimised for SST_{skin} measurements for satellite validation and while other instrument configurations are possible, the increase in integration time would be suboptimal for satellite validation.

[2]The ISAR project is funded since 2004 by Department for Environment, Food and Rural Affairs (Defra) (now Department of Energy and Climate Change (DECC)) to provide *in situ* SST_{skin} data for the AATSR validation

Chapter 5

Uncertainty analysis of *SST* measurements

5.1 Introduction

In this section the uncertainties of *SST* measurements are investigated. In particular the focus is to examine differences between SST_{skin} as measured by both satellite and shipborne infra-red radiometry, and SST_{depth} measurements from a ship's hull mounted thermometer. A number of methods were used to estimate the uncertainties of each measurement in order not only to obtain a better understanding of the uncertainties themselves but also to analyse the merits of the different uncertainty analysis methods. Three different methods were used to analyse the uncertainties of AATSR data (the satellite data), ISAR (the *in situ* SST_{skin} data set) and SBE 48 (the SST_{depth} measurements). The three uncertainty estimation methods used were the O'Carroll *et al.* (2008) uncertainty estimation (OC), the Tokmakian and Challenor (1999) mean square error estimation (TCM) and the Challenor (2004) uncertainty estimation (CH). The main difference between OC and TCM is that OC attempts to separate the bias and the random components, whereas TCM has a combined uncertainty for bias and random components.

5.2 Three way uncertainty analysis

Three way uncertainty analysis is a means to estimate the uncertainty of each data set used in the analysis. Two methods are discussed in this section, OC demonstrated on *SST* data and TCM demonstrated on sea surface height (SSH) data.

5.2.1 The OC method

The method described in this section was developed by O'Carroll *et al.* (2008) and further details about the method can be found in the mentioned paper.

The uncertainty in observation x_i, where i represents the different observation methods, can be expressed as

$$x_i = x_T + b_i + e_i \tag{5.1}$$

where x_T, is the true value of variable x, b_i is the bias (mean error) in the observation, and e_i, is the random error in the observation (which, by definition, has zero mean but may be non-Gaussian).

For a set of three collocated observations, $i = 1, 2$, and 3, we can write the following corresponding set of equations for their uncertainty:

$$\begin{aligned} x_1 &= x_T + b_1 + e_1 \\ x_2 &= x_T + b_2 + e_2 \\ x_3 &= x_T + b_3 + e_3 \end{aligned} \tag{5.2}$$

Now consider each set of three observations as three sets of pairs. The difference between observations i and j is given by

$$x_i - x_j = b_i - b_j + e_i - e_j \tag{5.3}$$

For an ensemble of such sets of observations, the mean difference between observations i and j is

$$b_{ij} = \bar{x_i} - \bar{x_j} = b_i - b_j \tag{5.4}$$

and the variance of the difference between these two observation types is

$$\begin{aligned} V_{ij} &= \left[(\bar{x_i} - \bar{x_j}) - (\bar{b_i} - \bar{b_j})\right]^2 = (\bar{e_i} - \bar{e_j})^2 \\ &= \bar{e_i}^2 + \bar{e_j}^2 - 2\bar{e_i}\bar{e_j} \end{aligned} \tag{5.5}$$

Therefore,

$$V_{ij} = \sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j \tag{5.6}$$

where σ_i^2 is the variance of the uncertainty in observation i, and r_{ij} is the correlation of uncertainty between the different observation methods i and j. Stating equation 5.6 explicitly for the following three sets of observation pairs:

$$\begin{aligned} V_{12} &= \sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2 \\ V_{23} &= \sigma_2^2 + \sigma_3^2 - 2r_{23}\sigma_2\sigma_3 \\ V_{31} &= \sigma_3^2 + \sigma_1^2 - 2r_{31}\sigma_3\sigma_1 \end{aligned} \tag{5.7}$$

The three simultaneous equation 5.7 can be solved to give the variance of uncertainty in each observation method as follows:

$$\begin{aligned} \sigma_1^2 &= \frac{1}{2}\left(V_{12} + V_{31} - V_{23}\right) \\ &\quad + \left(r_{12}\sigma_1\sigma_2 + r_{31}\sigma_3\sigma_1 - r_{23}\sigma_2\sigma_3\right) \\ \sigma_2^2 &= \frac{1}{2}\left(V_{23} + V_{12} - V_{31}\right) \\ &\quad + \left(r_{23}\sigma_2\sigma_3 + r_{12}\sigma_1\sigma_2 - r_{31}\sigma_3\sigma_1\right) \\ \sigma_3^2 &= \frac{1}{2}\left(V_{31} + V_{23} - V_{12}\right) \\ &\quad + \left(r_{31}\sigma_3\sigma_1 + r_{23}\sigma_2\sigma_3 - r_{12}\sigma_1\sigma_2\right) \end{aligned} \tag{5.8}$$

If the uncertainty in the three observation methods are uncorrelated, then $r_{ij} = 0$ for all $i \neq j$. In that case, equation 5.8 becomes

$$\begin{aligned} \sigma_1^2 &= \frac{1}{2}(V_{12} + V_{31} - V_{23}) \\ \sigma_2^2 &= \frac{1}{2}(V_{23} + V_{12} - V_{31}) \\ \sigma_3^2 &= \frac{1}{2}(V_{31} + V_{23} - V_{12}) \end{aligned} \tag{5.9}$$

This allows us to estimate the uncertainty variance in the three different observations methods from the observation difference statistics. However, the validity and accuracy of this method depends crucially on the assumption of the independence of uncertainty. The independence of the data and associated uncertainty is assessed in section 5.4.

5.2.2 The TCM method

The uncertainty estimation method described in this section was developed by Tokmakian and Challenor (1999) for the joint estimation of model and satellite sea surface height anomaly errors.

As before (O'Carroll *et al.*, 2008) we can define the three measurements as

$$\begin{aligned} x_1 &= x_T + e_1 \\ x_2 &= x_T + e_2 \\ x_3 &= x_T + e_3 \end{aligned} \tag{5.10}$$

where $x_{1,2,3}$ represent the different observation methods, x_T is the true value of the measured property (*SST*), $e_{1,2,3}$ the uncertainty associated with each measurement. It must be noted that it is assumed that the use of expectation operators for the true *SST* is well defined. Furthermore it is assumed that the bias in each measurement is small compared to the amount of variability in the measured property. If $E(x_{1,2,3}^2)$ are defined as mean square error (MSE), equation 5.10 can be rewritten:

$$\begin{aligned} E(x_1^2) &= E(x_T^2) + 2E(e_1 x_T) + E(e_1^2) \\ E(x_2^2) &= E(x_T^2) + 2E(e_2 x_T) + E(e_2^2) \\ E(x_3^2) &= E(x_T^2) + 2E(e_3 x_T) + E(e_3^2) \end{aligned} \tag{5.11}$$

$E(x_1^2)$, $E(x_2^2)$ and $E(x_3^2)$ are the unknowns which we are trying to determine. The expected values of the cross-products (covariances) are given by

$$\begin{aligned} E(x_2 x_1) &= E(x_T^2) + E(x_T e_2) + E(x_T e_1) + E(e_2 e_1) \\ E(x_3 x_1) &= E(x_T^2) + E(x_T e_3) + E(x_T e_1) + E(e_3 e_1) \\ E(x_2 x_3) &= E(x_T^2) + E(x_T e_2) + E(x_T e_3) + E(e_2 e_3) \end{aligned} \tag{5.12}$$

If the uncertainty of each measurement is independent of each other $E(e_2 e_1)$, $E(e_3 e_1)$ and $E(e_2 e_3)$ can be eliminated. Using equation 5.11 and 5.12 to eliminate the true value x_T and its covariance with error terms ($E(x_T^2)$, $E(x_T e_1)$, $E(x_T e_2)$ and $E(x_T e_3)$ the error variance can now be written in terms which can be evaluated directly from the measurements, as:

$$\begin{aligned} E(e_1^2) &= E(x_1^2) - E(x_2 x_1) + E(x_2 x_3) - E(x_1 x_3) \\ E(e_2^2) &= E(x_2^2) - E(x_2 x_1) - E(x_2 x_3) + E(x_1 x_3) \\ E(e_3^2) &= E(x_3^2) + E(x_2 x_1) - E(x_2 x_3) - E(x_1 x_3) \end{aligned} \tag{5.13}$$

5.3 A model for estimating the uncertainty

The method described in this section has been developed by P. Challenor for the calibration of the Radar Altimeter on Envisat (RA-2) (Challenor, 2004). While this method was developed for RA-2 calibration it can be used for any multi-way analysis. Also while the model was designed for instrument calibration using data sources of similar uncertainties, it is used here to calculate the variances of each of the data not to calibrate any of the data.

5.3.1 A model for calibration

To set up our model the assumption is made that m instruments all making simultaneous measurements of the same parameter, x, each have a sample population of size n. Assuming that all the desired calibrations are linear then the measured data y_{ji} (where j refers to the j^{th} instrument and i to the i^{th} data point) can be expressed as

$$y_{ij} = \alpha_j + \beta_j x_i + e_{ji} \tag{5.14}$$

α_j and β_j are the calibration constants for the j^{th} instrument, x_i is the true value of x at data point i and the errors e_{ji} will be assumed to be uncorrelated both between and within instruments and to be distributed $N(0, \sigma_j^2)$. The only measured variable in this equation is y_{ij}. The $\alpha's$, $\beta's$ and xs all have to be estimated. It is fairly easy to compute the Maximum likelihood estimator (MLE) for each of these parameters. They are

$$\hat{\alpha_j} = \frac{1}{n}\left[\sum_{i=1}^{n} y_{ji} - \hat{\beta_j}\sum_{i=1}^{n}\hat{x_i}\right] \qquad j = 1 \ldots m \tag{5.15}$$

$$\hat{\beta_j} = \frac{\left[\sum_{i=1}^{n} y_{ji}x_i - \hat{\alpha_j}\sum_{i=1}^{n}\hat{x_i}\right]}{\sum_{i=1}^{n}\hat{x_i}^2} \qquad j = 1 \ldots m \tag{5.16}$$

$$\hat{\sigma_j} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{ji} - \hat{\alpha_j} - \hat{\beta_j}\hat{x_i}\right)^2} \qquad j = 1 \ldots m \tag{5.17}$$

$$\hat{x_i} = \frac{\sum_{j=1}^{m}\frac{y_{ji}\hat{\beta_j}}{\hat{\sigma_j}^2} - \sum_{j=1}^{m}\frac{\hat{\alpha_j}\hat{\beta_j}}{\hat{\sigma_j}^2}}{\sum_{j=1}^{m}\frac{\hat{\beta_j}^2}{\hat{\sigma_j}^2}} \qquad i = 1 \ldots n \tag{5.18}$$

The first three of this set of equations are the standard results for MLE in regression if x is known. Therefore the MLEs can be computed by making an initial guess for α_j, β_j and σ_j $(j = 1, m)$, calculate x_i from equation 5.18, use a standard regression

package to estimate new values for α_j, β_j and σ_j for each j separately and iterate until the parameters converge.

5.3.2 Indeterminacy

Considering the original model in equation 5.14 and replace x_i with a linear transformation

$$x_i = a + bz_i \tag{5.19}$$

then equation 5.14 becomes

$$y_{ij} = \alpha_j + \beta_j (a + bz_i) + e_{ji} \tag{5.20}$$

$$y_{ij} = (\alpha_j + \beta_j a) + (\beta_j b) z_i + e_{ji} \tag{5.21}$$

Thus the same form of the equation is achieved when any linear transformation is applied to the true values of x. This means that α_j and β_j can not be determined absolutely. If all the instruments read too high by the same amount this method is not able to reveal that, and similarly if they all measure a change of two units when the real change is one. In order to produce a unique solution an extra constraint must be imposed. By using the constraint $\alpha_1 = 0$ and $\beta_1 = 1$ all the calibrations are relative to the first instrument. It should be noted that this does not imply that the first instrument is being used as a standard. σ_1 is still estimated. An alternative, and probably superior way of addressing the problem of indeterminacy would be to use a Bayesian approach and take into account the prior knowledge of the calibrations and true values of x. The equations for the MLEs are modified substituting 0 and 1 for $\alpha_1 = 0$ and $\beta_1 = 1$ where appropriate.

5.3.3 Variances of estimators

The variances of the estimators are derived from the second derivative of the log likelihood function. These give

$$var\left(\hat{\alpha_j}\right) = \frac{\sigma_j^2}{n} \qquad j = 2 \ldots m \tag{5.22}$$

$$var\left(\hat{\beta_j}\right) = \frac{\sigma_j^2}{\sum_{i=1}^{n} x_i^2} \qquad j = 2 \ldots m \tag{5.23}$$

$$var\left(\hat{\sigma_j}\right) = \left[\frac{3}{\sigma_j^4}\sum_{i=1}^{n}\left(y_{ji} - (\alpha_j + \beta_j x_i)\right)^2 - \frac{n}{\sigma_j^2}\right]^{-1} \qquad j = 2 \ldots m \tag{5.24}$$

$$var\left(\hat{\sigma_1}\right) = \left[\frac{3}{\sigma_1^4}\sum_{i=1}^{n}\left(y_{1i} - x_i\right)^2 - \frac{n}{\sigma_1^2}\right]^{-1} \tag{5.25}$$

$$var\left(\hat{x_i}\right) = \frac{1}{\left(\frac{1}{\sigma_1^2} + \sum_{j=2}^{m}\frac{\beta_j^2}{\sigma_j^2}\right)} \qquad i = 1 \ldots m \tag{5.26}$$

The estimation of the variances is achieved by replacing the parameters by their MLEs. In this case the variance of σ_j simplifies to

$$var\left(\hat{\sigma_j}\right) = \left(\frac{\sigma_j^2}{2n}\right) \qquad j = 1 \ldots m \tag{5.27}$$

5.3.4 A practical problem

Challenor (2004) found that when this method is implemented, a problem becomes apparent. The expectation maximization algorithm (EM) makes the x_T (the true values of *SST*) equal to the measured values for the instrument with the smallest error. To produce estimates in line with the a priori knowledge, a generalisation of the E-step (The

EM algorithm has two steps an E, or expectation step, and a M, or maximization step. The E-step computes the expected log likelihood function where the expectation is taken with respect to the computed conditional distribution of the latent variables given the current settings and observed data and M-step finds the parameters that maximises the log likelihood function.) is investigated. Rather than normalising by the error variance (σ^2), normalise by σ^γ. Experimentation shows that as $\gamma \to 2$ and $\gamma \to 0$ the results become unrealistic; x becoming equal to the data for $\gamma = 2$ and a constant for $\gamma = 0$. However around $\gamma = 1$ consistent results that are in line with our expectations are produced. These are used in the rest of this section.

5.4 Data

A few considerations had to be taken into account in respect of the data used for the three way uncertainty analysis. These considerations are that (a) the data have to be independent and not be correlated with each other, (b) the data had to be available, and (c) to make the results comparable with other studies it was important to use similar data sets. As shown in chapter 3 a number of *SST* measurements were recorded on the PoB, which makes the collocation of the datasets easy. After consideration of the above mentioned points it was decided to use the ISAR data, the hull contact thermometer data SBE 48 (see section 3.4.1.1 for details) and the AATSR data for the analysis. The choice of the ISAR data was not only to show the benefit of radiometer measurement for satellite validation but also to get an independent assessment of the uncertainty model shown in chapter 4.

For the three-way analysis four years (2005-2008) of ISAR, SBE 48 and AATSR data were used. While ISAR and AATSR data were used without any further processing and as described in chapter 3, the SBE 48 needed some processing before it could be used. Firstly the SBE 48 needed to be corrected for the warming which the hot environment of the PoB pump room introduced to the measurements and secondly it needed to be adjusted for the skin effect to be comparable to the ISAR and AATSR measurements.

The skin effect adjustment is necessary because ISAR and AATSR measure SST_{skin} and the SBE 48 measures SST_{depth} at approximately 5 m.

First the PoB pump room warming correction was carried out. The necessary bias correction was determined by comparing the SBE 48 data to CPR temperature data . The CPR temperature data is measured by a RBR 1050 temperature recorder attached to the CPR. The CPR was deployed once a month on the PoB on the northward leg (see section 3.4.1.3 for details). The CPR temperature recorder was deployed at a similar depth as the SBE 48 measurement of approximatively 6 m. To minimise effects such as diurnal warming and the ship's down-draw around the hull on the SBE 48 bias correction, only night time observations at high wind speeds were used (Wimmer and Robinson, 2008c). The data used for the SBE 48 bias calculation is shown in table 5.1. For the bias calculation night time was assumed to be 2.5 h either side of midnight and high wind conditions to be winds over a speed of $6ms^{-1}$.

TABLE 5.1: CPR - SBE 48 bias correction data. The label D marks the deployment number (see table 3.7) and date is shown in day of year. A warm sky brightness temperature shows a cloudy sky and a cold sky brightness temperature a clear sky.

D	Date	Bias (K)	Sd (K)	Wind Speed (m / s)	Sky temp	Data points
6	172.9 173.1 (2005)	0.333	0.0914	10 to 0	No data	288
7	259.9 260.1 (2005)	0.298	0.0884	> 10	No data	288
13	346.9 347.1 (2006)	0.247	0.0404	< 5	Cold / warm	288
14	73.9 74.1 (2007)	0.360	0.0435	No data	Cold	288
14	94.9 95.1 (2007)	0.597	0.0361	5 -10	Cold	288
16	193.9 194.1 (2007)	0.595	0.2032	≈ 5	Warm	288
17	286.9 287.1 (2007)	0.363	0.0904	< 5	Cold	288
17	317.9 318.1 (2007)	0.346	0.0887	> 10	Warm	288

Applying these considerations to table 5.1 results in two measurements usable for the bias correction. These are the D7 and the second D17 record. This led to a bias correction value of 0.332 K, which has been applied to the SBE 48 measurements for this comparison.

The second adjustment to the SBE 48 measurements was the correction for the skin effect. Here the approach of (Donlon *et al.*, 2002) was followed and a fixed offset for the skin effect of -0.17 K was applied. Temperature comparison between ISAR SST_{skin}

TABLE 5.2: Data labelling used throughout this chapter

Sensor	Label
AATSR	1
ISAR	2
SBE 48	3
AATSR - ISAR	12
ISAR - SBE 48	23
SBE 48 - AATSR	31

measurements and the SBE 48 SST_{depth} measurements showed a SST_{skin} to SST_{depth} difference of -0.15 K which is in line with the theoretical value of -0.17 K for the cool skin of the ocean (Donlon *et al.*, 2002; Soloviev and Schluessel, 1994). However to keep the datasets independent a fixed value for SST_{depth} to SST_{skin} conversion of -0.17 K was used.

Table 5.2 shows the labels used to distinguish the different datasets in the three way uncertainty analysis. For AATSR the label 1 is used for all four AATSR data products (D3, D2, N3, N2) with appropriately labelled captions.

5.4.1 Data quality and correlation

The data used for the uncertainty analysis has been produced by matching ISAR and SBE 48 data to AATSR data that was within 1 km and $\pm 2h$ (grade 2b in Wimmer *et al.* (2012)). The data quality of the three matched data sets was assessed by calculating the differences between the data sets and looking at the histograms and quantile-quantile plot (Q-Q plot) of the data. Figure 5.1 shows the histogram for AATSR D3 data and it is quite clear from the histogram that the data has some outliers which potentially need filtering.

To assess which filter limits would work best the Q-Q plot of the data were produced (see figure 5.2 and 5.3). A Q-Q plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.

The plots for the AATSR dual view two channel (day) product (D2), AATSR nadir view three channel (night) product (N3) AATSR nadir view two channel (day) product (N2)

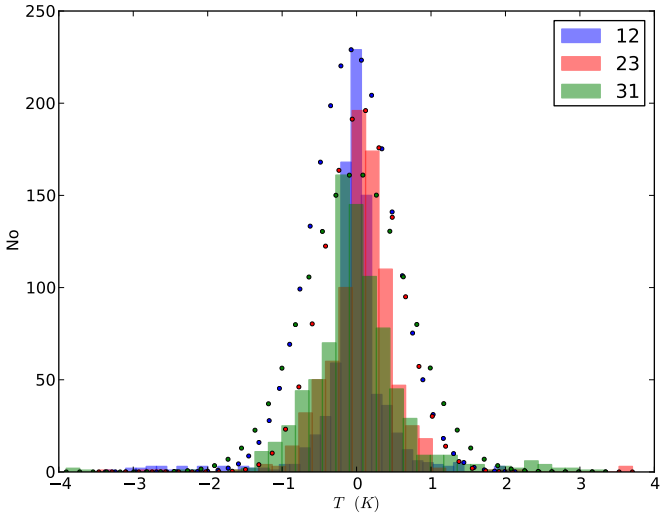


FIGURE 5.1: Histogram of the data difference for AATSR D3 data.

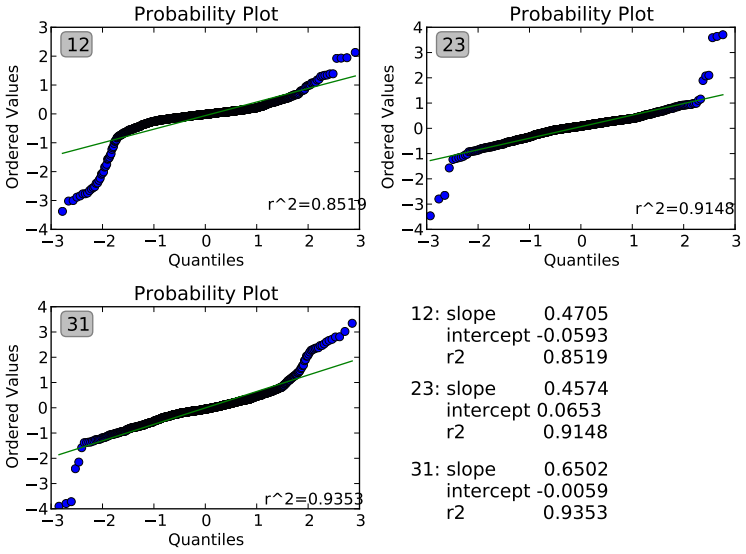


FIGURE 5.2: QQ-plot of the data difference for AATSR D3 data.

data can be found in the appendix A. From the Q-Q plot it was decided to use a one standard deviation filter on the data and Figure 5.3 shows the resulting Q-Q plot for the AATSR dual view three channel (night) product (D3) matched data. Table 5.3 shows the mean and standard deviation of the data differences after the data were filtered for the matched data corresponding to all four AATSR data products. The values are all very well controlled apart from the mean differences of the N2 product (12 and 31 are the N2 - ISAR and SBE 48 - N2 differences). This is due to a bias in the N2 product which has been documented by Corlett *et al.* (2006) and Wimmer *et al.* (2012). However

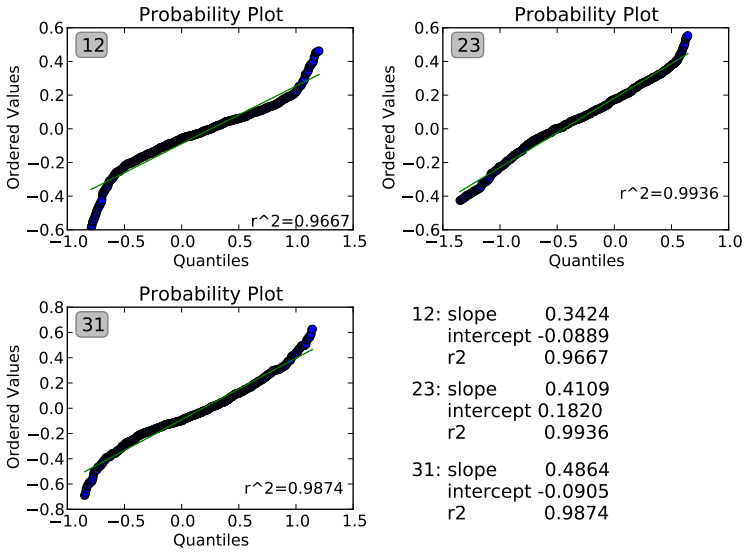


FIGURE 5.3: QQ-plot of the data difference for AATSR D3 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot.

TABLE 5.3: The mean and standard deviation of the differences for all four AATSR data products.

AATSR product	Label	Mean	*std*
D3	12	-0.03	0.19
	23	0.08	0.221
	31	-0.04	0.265
D2	12	-0.05	0.232
	23	0.05	0.262
	31	0.00	0.298
N3	12	0.06	0.14
	23	0.08	0.221
	31	-0.14	0.219
N2	12	0.62	0.245
	23	0.05	0.262
	31	-0.67	0.298

while the OC, TCM and CH uncertainty estimation methods demand a very low bias, we did not correct for the N2 bias. Tests with the SBE 48 bias have shown that the effect of a bias in the order of the N2 bias has very little effect on the results.

To be able to reduce the cross products in any of the discussed methods, the data used have to be independent of each other. This should be the case for this analysis where different sensor technologies were used for the calculation and the data used for SBE 48 calibration is derived from a measurement not used in this three way analysis. Also the

different sampling techniques have to be considered to meet the assumption ($r_{i,j} = 0$) required to derive equation 5.9. AATSR uses a near instantaneous spatial average, while ISAR is a temporal average over approximately 3 minutes of data and SBE 48 a one minute average re-sampled to the ISAR sampling. However from Wimmer *et al.* (2012) it has been shown that the ISAR three minute observation cycle covers a similar area as the AATSR $1km$ pixel while the PoB is travelling at 20 knots. Therefore the assumptions made for equation 5.9 should be met under most circumstances.

5.5 Results

Table 5.4 show the results for the three different three way uncertainty calculations OC (section 5.2.1), TCM (section 5.2.2, converted to a standard deviation) and the error calibration CH (section 5.3) for data collected between 2005 and 2008.

TABLE 5.4: Three way uncertainty analysis results for AATSR, ISAR and SBE 48 (labelled as SST_{depth}) for the years 2005 to 2008. Columns 3,4 and 5 show the sensors uncertainty.

Product	Method	AATSR	ISAR	SST_{depth}	Number	Unit
D3	OC	0.17	0.09	0.20	563	K
	TCM	0.17	0.09	0.20	563	K
	CH	0.15	0.10	0.19	563	K
D2	OC	0.19	0.13	0.23	326	K
	TCM	0.19	0.13	0.23	326	K
	CH	0.16	0.09	0.22	326	K
N3	OC	0.10	0.10	0.20	563	K
	TCM	0.10	0.10	0.20	563	K
	CH	0.04	0.10	0.19	563	K
N2	OC	0.20	0.14	0.22	326	K
	TCM	0.20	0.14	0.22	326	K
	CH	0.18	0.09	0.22	326	K

The results in table 5.4 show that the night time uncertainties (D3 and N3) are lower than the day time uncertainties (D2 and N2) for all sensors. They also show that for AATSR data N3 has the lowest uncertainty followed by D3 uncertainty and the two day time AATSR products D2 and N2 which show as similar uncertainty. These results confirm that AATSR three channel retrievals (N3, D3) perform better than the two

channel retrievals (N2, D2), though surprisingly there is very little difference between the N2 and D2 results. AATSR validation studies such as Wimmer *et al.* (2012) and Corlett *et al.* (2006) showed that the D2 outperforms the N2 product, but this does not seem to emerge from this analysis. This might be due to the fact that, while the AATSR dual view is generally more robust against atmospheric changes, it also introduces the uncertainty from the forward view channels into the equation therefore cancelling out some of the performance advantages the dual view channels show in classic validations. The ISAR uncertainties are the lowest of all three sensors, and the SBE 48 uncertainties are the highest in these comparisons, with the AATSR uncertainties between the ISAR and SBE 48 uncertainties.

The results are similar to O'Carroll *et al.* (2008) which estimates the uncertainty for D3 10 arc minute average of 0.12 to 0.16 K, for Advanced Microwave Scanning Radiometer - EOS (AMSR-E) 0.42 to 0.51 K and for drifting buoys 0.22 to 0.27 K. The results also tie in with those of Beggs *et al.* (2012) who claim that SBE 48 have a similar uncertainty to drifting buoys. The results also show that OC and TCM produce very similar results which is to be expected as these are very similar methods, with the only difference being the estimation of the systematic component of the uncertainty. It is interesting that CH produces a slightly lower uncertainty in all but the D3 ISAR case.

5.6 Discussion

Three-way or multi-way uncertainty analysis give an independent way of assessing the uncertainties associated with each data set. Three different methods were assessed in this chapter, with OC and TCM being very similar, which gives near matching results for OC and TCM. However while the results are near identical the main difference between OC and TCM is that TCM only requires independent data, while OC requires a few more assumptions about the data to be met. The assumptions for OC, such as near zero bias and similar space-time difference, are all met for the data used in this chapter although this might not always be the case. The near zero bias requirement for the OC method seems to be in contradiction to equations 5.2 - 5.4, where the bias is shown to

be estimated by the OC method. However the bias and random component can not be separated and therefore the bias has to be assumed to be near zero in order to get results for the random component from the OC method (O'Carroll *et al.*, 2008). This is a weakness of the OC method and together with the fewer assumptions needed for the TCM method it can be said that the TCM method is the more robust method for general use. The third method (CH) gives the smallest uncertainties for all the measurements which is a design intention of this method, as it minimises the uncertainty. However there is a small issue with the MLE estimation of the parameter γ which has not been fully resolved. Challenor (2013) is exploring the issue, however to date there has not been a satisfactory solution to estimate γ in a deterministic way.

Finally the uncertainty values computed with the three methods are comparable to O'Carroll *et al.* (2008), with O'Carroll *et al.* (2008) showing slightly lower uncertainties for the AATSR D3 10 arc minute product than the results shown here for the AATSR 1 km product. This is not surprising as the AATSR 10 arc minute product averages over a number of $1km$ pixels which should lower the uncertainty. The results for SBE 48 are similar to the ones for drifting buoy data and seem to confirm the point Beggs *et al.* (2012) makes that SBE 48 are a very good complement to drifting buoy data with similar uncertainties. The uncertainty analysis in this chapter also confirm that ISAR has the lowest uncertainty in all but one case, the N3 data.

The methods described in this chapter allow one to independently assess the uncertainties of the satellite derived SST_{skin}, the *in situ* SST_{skin} measurement and the *in situ* SST_{depth} measurement. The results in table 5.4 show that as long as the climate stability requirement of $0.04K\ decade^{-1}$ (Ohring *et al.*, 2005) is met satellite SST_{skin} has a lower uncertainty than SST_{depth} and therefore has the potential to be a better source for a CDR. The results also shows why it is important to use *in situ* radiometers for the validation of satellite SST as the low uncertainty of *in situ* radiometers compared to other *in situ* measurements will reduce the uncertainty they introduce to the validation process. The next chapter (6) looks into the validation process in more detail. A more detailed discussion on how the results of this chapter tie into the main objectives of this thesis can be found in chapter 7.

Chapter 6

Uncertainty analysis of the $match-up$ process

6.1 Introduction

In the previous two chapters the uncertainty of the *in situ* validation data (chapter 4) as well as the uncertainty of the AATSR measurement (chapter 5) were investigated. This chapter focusses on the process by which independent *in situ* SST_{skin} and satellite SST_{skin} datasets are compared in order to validate one of them. It explores the assumptions and approximations inherent in the conventional matching of "coincident" pairs of SST data acquired from different sources in order to produce a validation dataset. It assesses the uncertainties that imprecise matching and related factors introduce into the validation process and develops a practical approach towards quantifying these uncertainties. This capability then provides the analytical tools with which to develop a new approach to satellite SST validation that seeks both to quantify and minimise the extent to which the mismatch between two datasets is attributable to the $match-up$ process rather than the errors or uncertainties in the individual datasets. The chapter will first describe the conventional $match-up$ method and then develop a method which estimates the variability of the ocean and how it affects the $match-up$ process and ultimately the

validation result. The method presented here was first described in preliminary form by Wimmer and Robinson (2008a).

6.2 Classic Match-ups

Geophysical data validation is a fairly simple process that uses, most commonly, a predefined $match-up$ window to match a sample from one dataset with a sample from a different dataset that is deemed to be coincident in space and in time within the limitations defined by the window. To design the $match-up$ window a time window and a distance window are required that will, on the one hand, maximise the number of matched data pairs for the validation and on the other hand minimise errors or uncertainties introduced when the sample from one dataset is too far removed in space or time from the matched sample in the other dataset. While a $match-up$ window that is too restrictive will not take advantage of all the possible data available, a $match-up$ window which is too wide is a bigger issue as it will compare data that are quite different, due to geographical or temporal separation.

Figure 6.1 shows four $match-up$ scenarios for generating validation data sets. The four scenarios can be divided into matching satellite data to a fixed location such as a moored buoy in the top row (a and b) and matching satellite data to data acquired along a track such as ship observations in the bottom row (c and d). Furthermore the figure shows ideal conditions in the left hand (a and c) and cloudy conditions in the right hand column (b and d).

Figure 6.1 (a) shows the ideal $match-up$ conditions between a satellite pixel and a fixed mooring. Because a particular satellite pixel covers the mooring position no (or to be precise a very small) spatial window is needed in the matching process and the temporal window can be very small depending on the buoy data recording interval which can vary from minutes to hours. Figure 6.1 (b) shows the same situation as (a) but with a cloud covering the satellite pixel closest to the buoy location. In this case a $match-up$ with a satellite pixel a certain distance from the buoy location (the search radius $N\Delta x$) can still produce a validation point. However in this case it has to be ensured that the part

of the ocean measured by the buoy and the part of the ocean measured by the satellite pixel are similar.

Figure 6.1 (c) shows the ideal $match - up$ conditions between a satellite pixel and a measurement moving along a track, such as an *in situ* measurement on a ship. The *in situ* sensor will be at the same location as the satellite pixel, however there might be a time difference between the satellite pixels acquisition and the measurement at the same location resulting in a wider temporal $match - up$ window than normally needed for a fixed buoy installation. This wider temporal window compared to a fixed location might be necessary because as the ship move along it might not be possible to find a match where the ship is in the same location at the same time as the satellite pixel was acquired. Figure 6.1 (d) shows the same situation as (c) but with a cloud covering the satellite pixel closest to the along track *in situ* measurement. In this case we not only need a spatial search window (like in the fixed buoy case) but also a temporal search window $M\Delta t$ to allow for the ships movement and therefore location change. Using a non ideal $match - up$ between the satellite pixel and the along track *in situ* measurement can result in a poor comparison, however if it can be demonstrated that the oceanographic and atmospheric conditions of the non ideal $match - up$ pair are equivalent this will still yield a good validation data point.

The classic $match - up$ method uses a temporal and spatial window to define where the *in situ* and satellite measurement are seen as coincident. Minnett (1991) concluded that for *SST* a $match - up$ window of ± 2 h and ± 10 km is usable in most conditions. Corlett *et al.* (2006) uses ± 2 h and ± 1 km, where the 1 km limit is meant to restrict the *in situ* measurement to be in the same satellite pixel as the AATSR measurement to which it is matched. Wimmer *et al.* (2012) uses five $match - up$ windows to investigate what effect the window has on the $match - up$ process. Table 6.1 shows the $match - up$ windows used in Wimmer *et al.* (2012).

The different $match - up$ windows, or $match - up$ grades, are a way of stratifying the data in order to establish the trade off between maximising the number of $match - up$ pairs and the quality of the $match - up$ pairs. However this method assumes that a close match in time and location has to be a good match, which is not always the case. For example,

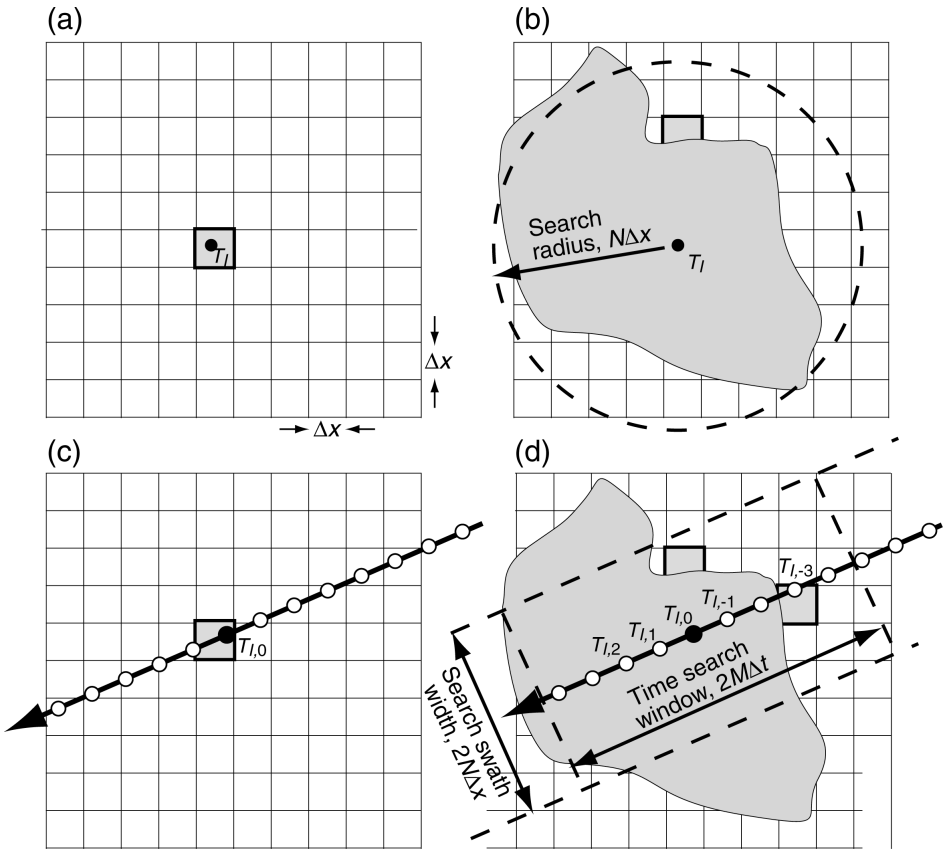


FIGURE 6.1: Examples of *match − up* situations encountered in the construction of a *match − up* database. (a) Point sample when there is no cloud. Match the *in situ* sample closest to the time of the overpass to the pixel in which it lies. (b) Point sample obscured by cloud. Match the *in situ* sample closest to the time of the overpass to the closest cloud-free pixel. The search radius needs to be limited to N pixels. (c) Along-track sensor such as ISAR in cloud free conditions. Match the *in situ* sample closest to the time of the overpass to the pixel in which it lies. (d) Along-track sensor in cloudy conditions. From Wimmer and Robinson (2008a)

TABLE 6.1: Match-up windows for AATSR validation

Grade	Temporal window (h)	Spatial window (deg)
1	0.5	0.01
2a	0.5	0.2
2b	2.0	0.01
3	2.0	0.2
4	6.0	0.225

in the vicinity of ocean fronts, which have a large temperature gradient, comparing the *SST* from one side of the front with *SST* of the other side of the front can result in a few Kelvin difference, giving a poor validation data point. Also some data might be excluded from the *match − up* process because the data are collected quite far apart, even though the *SST* variability in the open ocean is small and would therefore lead to

a good comparison.

Figure 6.2 shows an example of an ideal *match − up* situation with the *match − up* pair being located in a cloud free area with a near uniform temperature. In contrast figure 6.3 shows a *match − up* situation which is far from ideal, not only is there cloud in the scene, there is also a considerable temperature gradient near the location of the *match − up* pair.

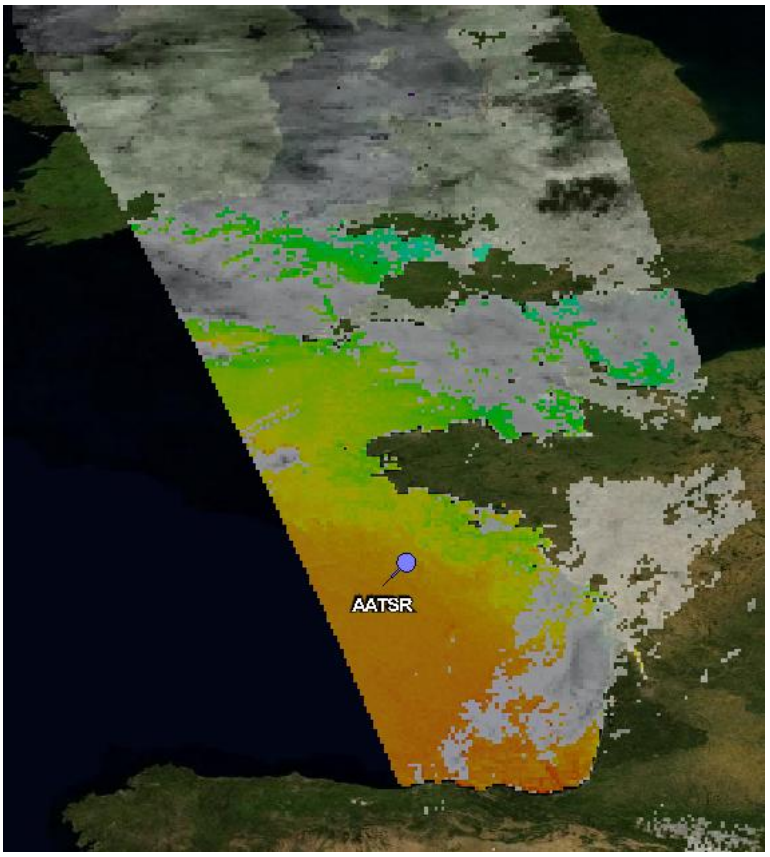


FIGURE 6.2: AATSR image from 16.03. 2009, showing a *match − up* scene where *match − up* windows work well. The AATSR pin shows the AATSR pixel matched to ISAR with a grade 2b *match − up* window. From Hopkins (2010).

6.3 The *match − up* uncertainty budget

To reduce the shortcomings that a fixed *match−up* window introduces in the comparison of two datasets the validation process was analysed from first principles. This means looking at what the validation process tries to achieve, a comparison of the satellite measured value (x_S) with a true value (x_T) and an estimate of how big the difference

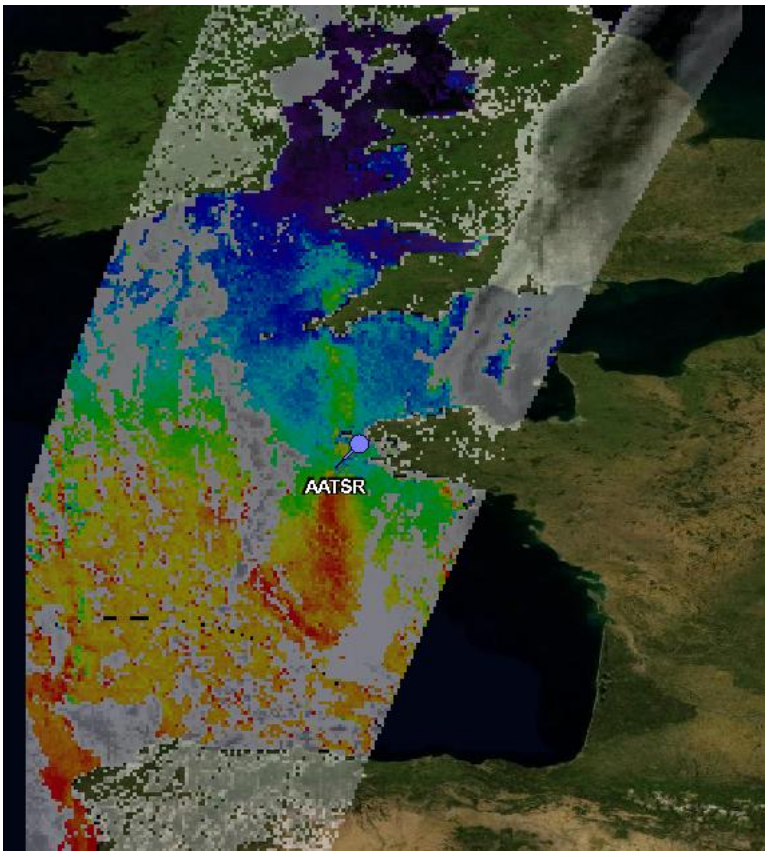


FIGURE 6.3: AATSR image from 12.04.2009, showing a *match − up* scene where *match − up* windows give a high difference between the *match − up* pair. The AATSR pin shows the AATSR pixel matched to ISAR with a grade 2b *match − up* window. From Hopkins (2010).

between the two is, which is called the error (e) between the two measurements. However, because the true value is not known, the satellite measurement has to be compared to another measurement, which can be an *in situ* measurement (x_I) or another satellite measurement. In this thesis only the comparison to an *in situ* measurement is considered. Also because the true value is not known, it is not possible to calculate the error of the satellite measurement, but the uncertainty u_M with which the true value is estimated. Equation 6.1 shows how the error (e) is calculated as the difference between the satellite-derived value (x_S) and the true value (x_T), assuming the true value is known.

$$e = x_S - x_T \tag{6.1}$$

Rewriting equation 6.1 to allow for not knowing the true value:

$$u_M = x_S - x_I \tag{6.2}$$

where u_M is what a classic validation would give as a result, which does not allow to assign the uncertainty u_M to either measurement (x_S or the *in situ* measurement, x_I). To be able to quantify the factors which contribute to the $match - up$ uncertainty and therefore get a better understanding of what fraction of the $match - up$ uncertainty can be assigned to the satellite measurement uncertainty the main components contributing to the $match - up$ uncertainty are discussed in the following paragraphs.

The spatial mismatch between the satellite measurement and the *in situ* measurement introduces an uncertainty component, which is due to the fact the the satellite measurement and the *in situ* measurement are made at geographically different parts of the ocean. This is not a large contribution to the measurement uncertainty in open ocean condition but can have a big impact on the uncertainty budget in frontal regions. This spatial mismatch uncertainty will be called u_S.

The temporal mismatch between satellite measurement and the *in situ* measurement introduces another uncertainty component, which can be similar in nature to the spatial mismatch uncertainty.

The next component to consider is the instrument uncertainty u_I of both the satellite instrument and the *in situ* instrument. Because the satellite instrument uncertainty is the uncertainty this $match - up$ process is trying to estimate, only the *in situ* instrument uncertainty is considered here. Chapter 4 discusses the ISAR uncertainty model in detail, however because the ISAR uncertainty model was not available when this particular study of $match - up$ uncertainties was carried out a slightly different approach was used. The ISAR uncertainty is fairly well controlled, the instrument being self calibrated and the CASOTS II showing when the ISAR uncertainty is not within the specification of $\pm 0.1K$. Therefore only data within specification was allowed for this process, implying $u_I < 0.1\ K$.

Next the different sampling technologies of the satellite measurement and the *in situ* measurement have to be considered. While the satellite measures a near instantaneous

spatial average, the *in situ* measurement is recorded as a temporal average of a moving or fixed point. The point-in-area sampling uncertainty (u_P) is the uncertainty associated with using a single point sample from the *in situ* measurement to represent the average of a variable over the whole instantaneous field of view of a satellite measurement.

And finally the sampling depth has to be considered. Comparing a surface measurement (e.g.: SST_{skin}) with a measurement recorded at depth (e.g.: SST_{depth}) will introduce an uncertainty into the overall $match-up$ uncertainty budget. This uncertainty will be called the sampling depth uncertainty u_Z.

Considering all the uncertainty components discussed above equation 6.2 can be rewritten to identify the sources of uncertainty as

$$u_M = u_S + u_T + u_I + u_P + u_Z \tag{6.3}$$

Having considered the main contributing factors of the $match-up$ uncertainty the magnitude of these uncertainties has to be estimated. However they can not easily be modelled or estimated because the detailed information about the exact circumstances of each match-up is not available. Therefore a first principles approach to estimate the uncertainties can not be followed. However it is possible, with the information available, to generate for each type (t) of uncertainty an indicator (I_t) which will allow us to characterise the expected magnitude of each of uncertainty t for the particular context of each $match-up$ pair.

A set of I_t for each of the uncertainties in equation 6.3 can be evaluated uniquely for each $match-up$ pair allowing the relative uncertainty of one $match-up$ to be compared with the uncertainty of another $match-up$ pair, but only with respect to the same type of uncertainty. Thus a higher value of an indicator may imply a greater (or lesser) magnitude of that particular uncertainty type but it carries no information about how the indicator of one uncertainty type relates to that of another type. Therefore it is not possible to estimate the overall $match-up$ uncertainty of a $match-up$ pair, simply by adding the individual indicator values for the different uncertainty types.

Because an analytical combination of the indicators into a $match - up$ uncertainty is not possible an empirical method was developed to produce a quality indicator rather than a quantitative uncertainty. A statistical multivariate analysis is used to determine a $match - up$ quality indicator (QI) for each match-up pair, based on the combination of indicator values (I_{t1}, I_{t2}, I_{t3}, ..) across the uncertainty types (t_1, t_2, t_3, ..). QI is an integer classification scale from 0 to 5, with higher values indicating reducing uncertainty attributable to the match-up process itself. The formulation of indicator combinations was evaluated as a correlation between the QI and the difference between x_S and x_I. The empirical confirmation for a suitable combination of indicator values is that if the match-up dataset is filtered by different values of QI (i.e. 3, 4, 5) the standard deviation of $x_S - x_I$ will reduce as QI increases.

6.4 The estimation of the $match - up$ uncertainty

The previous section outlined an approach for classifying the $match - up$ uncertainty by estimating each of the contributing uncertainties. However by analysing the components of equation 6.3 two uncertainties can be eliminated by using ISAR measurements. Because ISAR is a SST_{skin} measurement just as AATSR the sampling depth uncertainty (u_Z) can be set to zero. Furthermore the measurement instrument uncertainty becomes very well controlled by using ISAR measurements and was therefore eliminated from the analysis of the remaining $match - up$ uncertainties although ignoring the ISAR uncertainty places the baseline for the satellite uncertainty at no less than ± 0.1 K.

Thus equation 6.3 can be reduced to:

$$u'_M = u_S + u_T + u_P \tag{6.4}$$

leaving the spatial mismatch uncertainty, the temporal mismatch uncertainty and the point in sampling uncertainty representing the $match - up$ uncertainty. The full analysis of these three remaining uncertainties is not possible, but by considering their nature it is possible to estimate the typical magnitude of each. These are taken to be the

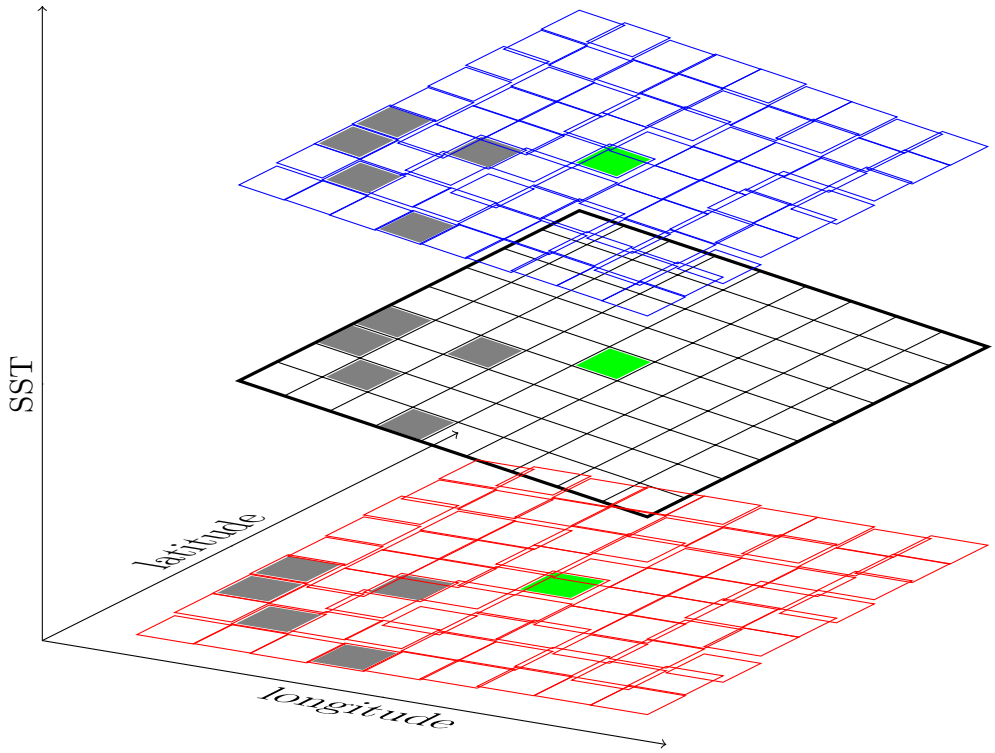


FIGURE 6.4: Schematics of the de-trending process. Shown in blue is the SST data as captured by AATSR, in black is the fitted flat plane and in red the de-trended SST data. The green pixel shows the $match-up$ pixel and grey pixels show cloudy data.

uncertainty indicator I_S, I_T and I_P. Each of the uncertainty indicators is calculated from the $match-up$ data set using objective criteria.

The point-in-area sampling uncertainty (u_P) depends on small scale variability of SST_{skin} in space and time, leading to apparently random differences between what ISAR measures along a line during the sea view part of ISAR sampling cycle and what AATSR averages over $1km^2$. To estimate u_P two different ways were explored. Firstly the variance of a 9 x 9 pixel field extracted from the AATSR image data and centred at the $match-up$ pixel was examined. To make sure only the small scale variability of SST_{skin} contributes to the variance the 9 x 9 pixel field was first de-trended by fitting a plane to the 9 x 9 field. The fitting was done by using a least square fit. The plane was then subtracted from the 9 x 9 data and the variance calculated for the remaining anomalies. This was done in order to only have the higher frequency components contributing to the variance calculation and avoid having a potential background gradient contaminating the small scale variability which this indicator tries to assess. Figure 6.4 shows a graphical representation of the de-trending process.

Because cloudy pixels would reduce the variance of the 9 x 9 *SST* field and lead to an underestimation of the small scale variability, the variance was weighted. The weighting factor was estimated by using the 95% confidence interval of a χ^2 distribution with one degree of freedom for the number of *SST* values available in a 9x9 *SST* field. The weighted variance is referred to as "explained variance". The "explained variance" can then be used as an indicator that characterises the uncertainty of using a single point to represent the average temperature over a whole pixel. This relies on the assumption that explained variance estimated over 9x9 *SST* field can be downscaled to the sub-pixel level length scales. As a variance, it corresponds to the typical squared magnitude of u_P. We refer to this indicator as I_{P1}:

$$I_{P1} = var(9x9)_{AATSR_{cloud\ free}} \cdot \chi^2_{95\%} \tag{6.5}$$

Another way of estimating u_P is to look at the other data source used in the $match - up$ process, the ISAR measurements. The ISAR data is recorded along the ship track and therefore, unlike in the AATSR data, a one dimensional data record over $\pm 10km$ either side of the $match - up$ was used. As before the data was first de-trended to remove any background gradient from the variance calculation, and the explained variance was again used. The reason for using the explained variance for ISAR data is that the instrument does not measure *SST* data when rain or sea spray would contaminate the optics of the instrument. The resulting $match - up$ indicator is referred to as I_{P2}. There might be an argument that I_{P1} and I_{P2} are to some extent correlated as they are estimated from the same *SST* structure, apart from the fact the ship takes about 35 minutes to cover the 20 km and thus far longer than the satellite measurement. However, for the purpose of setting quality values we consider them as separate indicators:

$$I_{P2} = var(ISAR)_{All\ samples\ 10km\ either\ side\ of\ match-up} \cdot \chi^2_{95\%} \tag{6.6}$$

The spatial mismatch uncertainty (u_S) occurs when the ISAR measurement of *SST* is for a location different from the satellite pixel to which it is matched. It depends on both the distance of the separation, Y, and the expected spatial gradients of *SST*, G_S,

in the region. This can be estimated from the satellite *SST* field smoothed to remove high frequency noise shorter than 5-10 km, which has already been taken into account through its contribution to the point sampling uncertainty (u_P). Then we can define another indicator:

$$I_S = G_S \cdot Y \tag{6.7}$$

The temporal mismatch uncertainty (u_T) is similarly the product of the temporal gradient of *SST*, G_T, and the time mismatch, t_{diff}, between the AATSR and the ISAR measurements. It is difficult to identify a measurable estimate of G_T that is fully independent of spatial variability, but we have compromised by using the time trend of *SST* measured by ISAR over a 1 hour window of data. Then the indicator for this uncertainty becomes:

$$I_T = G_T \cdot t_{diff} \tag{6.8}$$

6.5 Design of a quality value

The aim of the indicators as described in the previous section is to generate a $match-up$ QI, based on the uncertainties in the $match-up$ process. As discussed in section 6.3 this intermediate step to estimate the $match-up$ uncertainty is necessary, because a direct estimation of the $match-up$ uncertainty is not possible. The following section will present how the indicators are used to estimate the quality of a $match-up$ pair. Furthermore to offer some form of discrimination of the QI, six quality levels are used for Q ranging from 0 for a non quality assessed $match-up$ pair to 3, 4 and 5 quality assessed $match-up$ pairs. Note that the $match-up$ quality levels 1 to 5 aim to correspond to the scale values used internationally as indicators of satellite *SST* data quality (Corlett, 2007) where 3 is intended to represent the qualitative norm of "acceptable", 4 is "good" and 5 is "excellent". In practice no attempt has been made in this work to use or set thresholds for levels 1 and 2 which correspond to unacceptable quality of $match-up$ and would therefore be of no value in validation work.

6.5.1 Estimating the QI thresholds

After deciding which indicators and what QI levels would be used, criteria for what constitutes a QI level 3, 4 and 5 were needed. Data from ISAR deployment D11 on the PoB were used as a training data set to evaluate each indicator. Equations 6.5 to 6.8 were used to estimate I_{P1}, I_{P2}, I_S and I_T. Each indicator was assigned its own QI value Q_{P1}, Q_{P2}, Q_S and Q_T and thresholds were derived for each indicator to translate each indicator into a QI. The thresholds for each indicator were derived in an iterative empirical process in the following way:

(i) A number of of $match-up$ scenes in the training data set were analysed to get an understanding of what effect the different indicators have on the overall $match-up$ quality. This initial analysis showed the need for an additional parameter to be included in the QI selection. One of the issues of any validation process is the atmospheric correction, and an important contribution to the uncertainty of the *SST* retrieval from Top of Atmosphere (TOA) radiances is undetected clouds.

To evaluate whether cloud contamination in the $match-up$ process is a possibility the BT_{sky} measurement of the ISAR instrument was used as a cloud indicator. The ISAR BT_{sky} measurement is made every ISAR measurement cycle for the ISAR *SST* processor atmospheric correlation. High BT_{sky} values (close to ambient temperature) show cloud cover and low BT_{sky} values show clear sky measurements. The BT_{sky} was therefore used as a cloud indicator I_{sky} to define an additional contributor, Q_{sky}, to the quality value.

(ii) The analysis resulted in a first set of thresholds which then were evaluated by checking their performance against different scenes from the training data set.

(iii) Next the thresholds were fine tuned for maximum performance measured by the amount of data points classified as $match-ups$ in non ideal conditions while not classifying $match-ups$ in ideal conditions as being of low quality.

(iv) Finally the derived thresholds were evaluated on a longer time series, the whole of the training data set of three months, before the thresholds were finalised.

Figures 6.5 to 6.8 show graphical plots of applying the finally selected thresholds to evaluate the quality indicators assigned to each of the uncertainty types and overall, for the $match-up$ data pairs retrieved from two separate overpasses of AATSR, one at night on 10 July 2006 (figures 6.5 and 6.6) and one in the daytime on 28 July 2006 (figures

In figures 6.5 and 6.7 each $match-up$ pair is plotted by the value of u_M, (the difference between the AATSR (dual view product) and the ISAR temperatures) on the y axis and the time difference of the ISAR record before (-ve) or after (+ve) the AATSR overpass, on the x axis. This plot is presented eight times, but with different colour dots depicting the assigned quality indicator as follows. The overall QI is shown in the top panel, the second panel shows the Q_{sky}, the third panel shows the Q_{P2}, the fourth panel shows a variant of Q_{P2}, which does not have the trend removed from the data, the fifth panel shows Q_{P1}, the sixth panel shows the wind speed which is for reference only and was not used in the QI design, the seventh panel shows Q_S and the eighth panel shows Q_T. In all but the wind speed panel red dots represent non quality assessed $match-up$ pair ($Q=0$), yellow dots represent $Q=3$ $match-up$ pairs, light green dots represent $Q=4$ $match-up$ pairs and dark green dots represent $Q=5$ $match-up$ pairs.

Figures 6.6 and 6.8 show the histograms for how the separate QI values for all the indicators and the overall QI are distributed across the possible values 0, 3, 4 and 5. Note that the overall QI distribution is shown in its own column whereas the others are stacked together. The top panel shows the histogram for the dual view $match-up$ pairs and the bottom panel for the nadir view $match-up$ pairs. The plots show the Q_{P2} in green, Q_T in blue, Q_{sky} in red, Q_S in magenta, Q_{P1} in yellow and the overall Q in black. The bars at the different quality levels show where each individual QI would have placed the $match-up$ pair, however because the lowest individual QI determines the overall Q a $match-up$ pair can be classed as $Q=5$ for all but one QI and still be $Q=0$ if that one QI is 0.

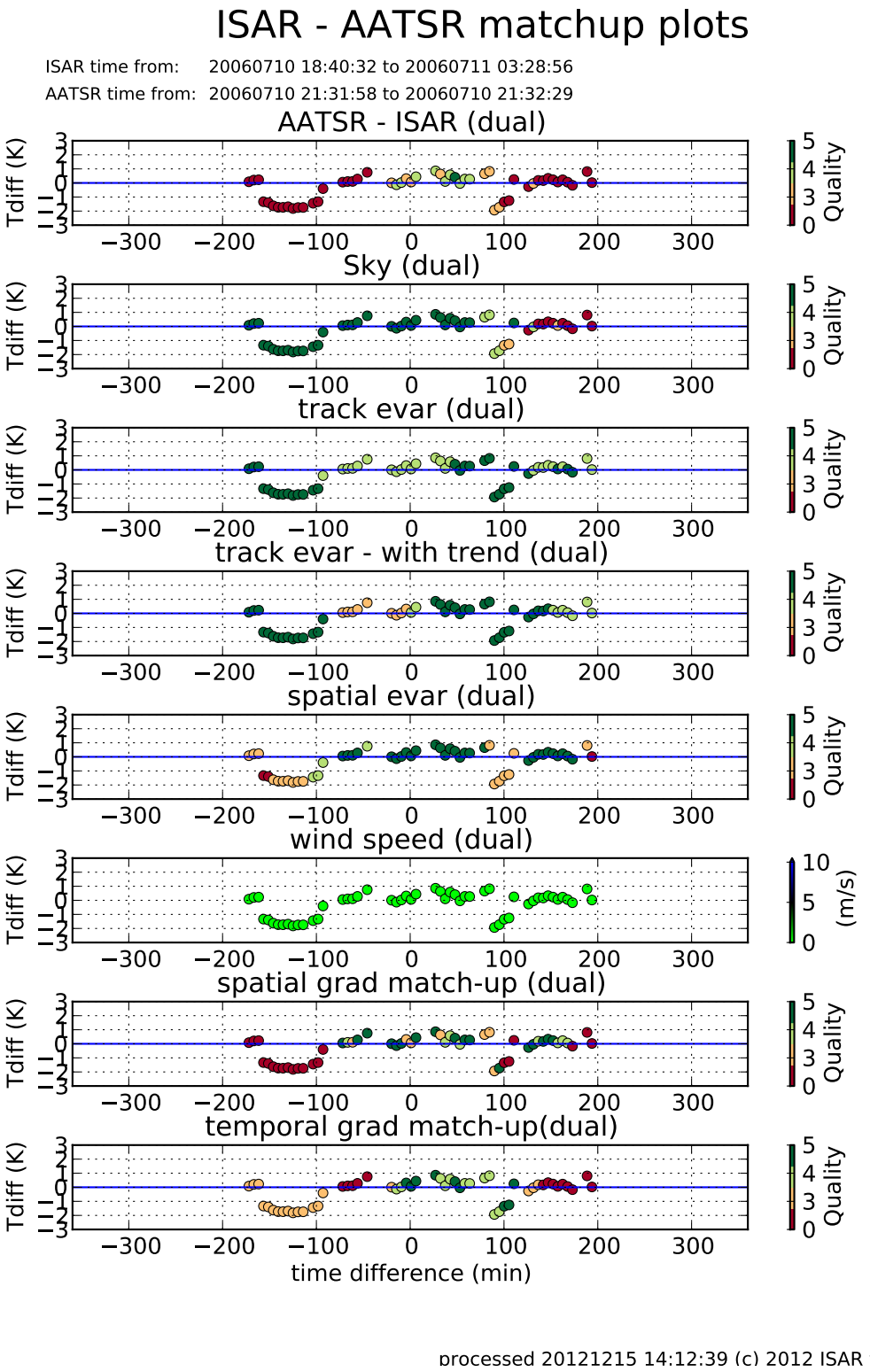


FIGURE 6.5: Quality indicator plot showing how the different parameters set the quality level for the *match − up* pair for the AATSR dual view night time *match − up* from the training data set. The the colour of the dots represents in all but the wind speed plot the QI, with dark green representing QI 5, light green QI 4, yellow QI 3 and red QI 0.

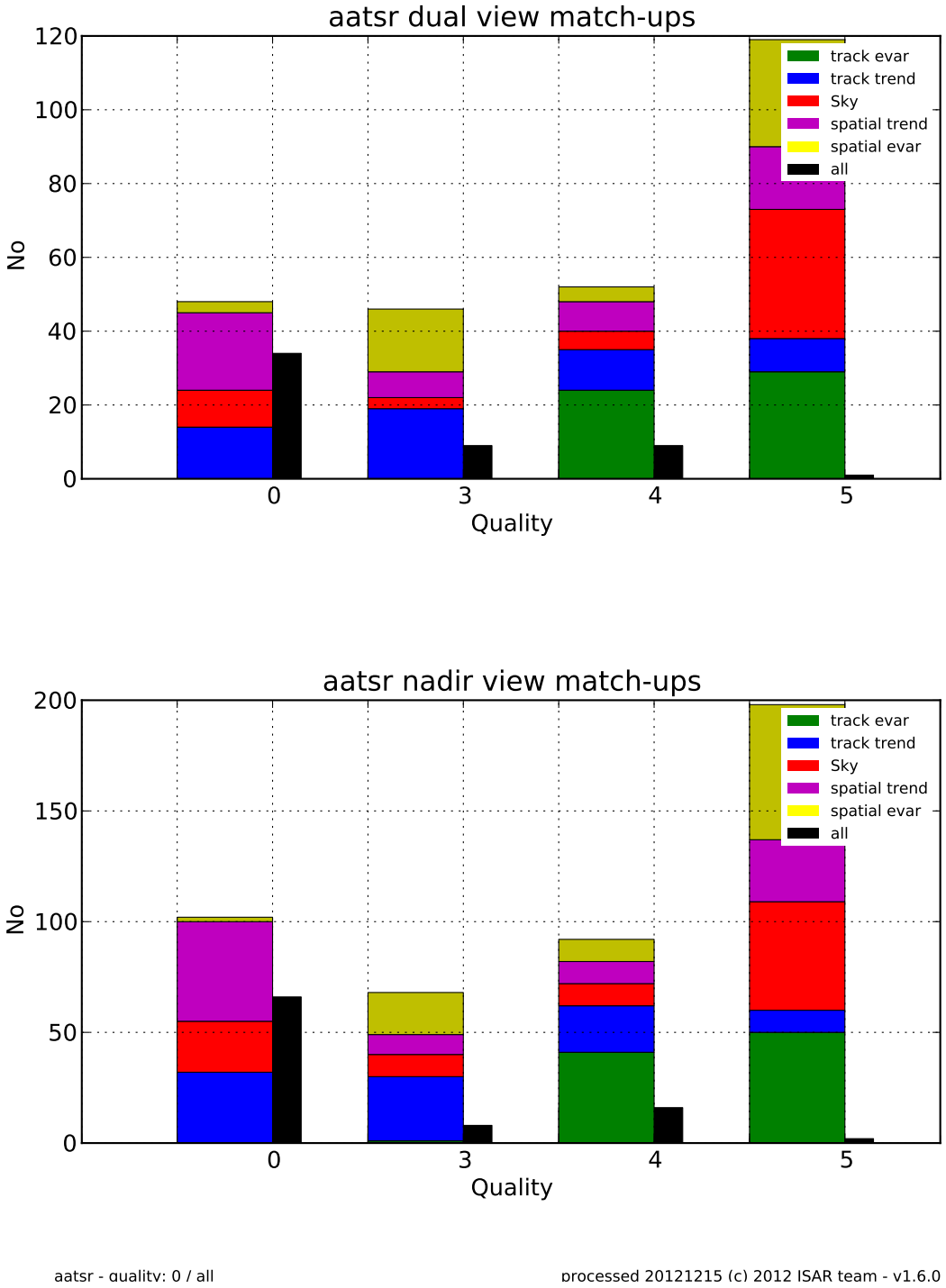


FIGURE 6.6: Quality indicator plot showing how the different parameters set the quality level for the $match - up$ pair for the AATSR night time $match - up$ from the training data set. Note that the bar plots were stacked on top of each other for better visibility and only show which indicator has the most impact in the given scene.

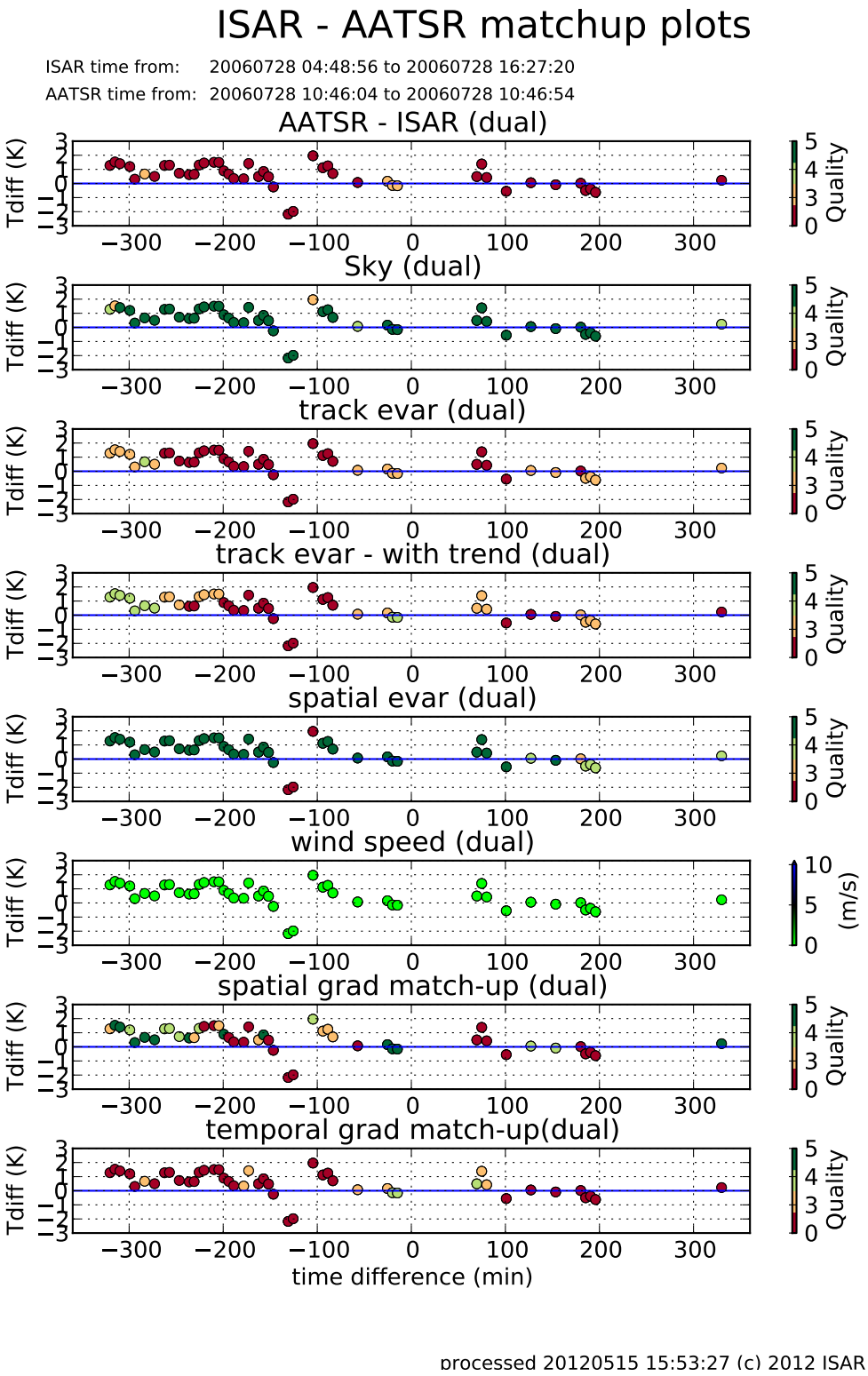


FIGURE 6.7: Quality indicator plot showing how the different parameters set the quality level for the *match − up* pair for the AATSR dual view day time *match − up* from the training data set. The the colour of the dots represents in all but the wind speed plot the QI, with dark green representing QI 5, light green QI 4, yellow QI 3 and red QI 0.

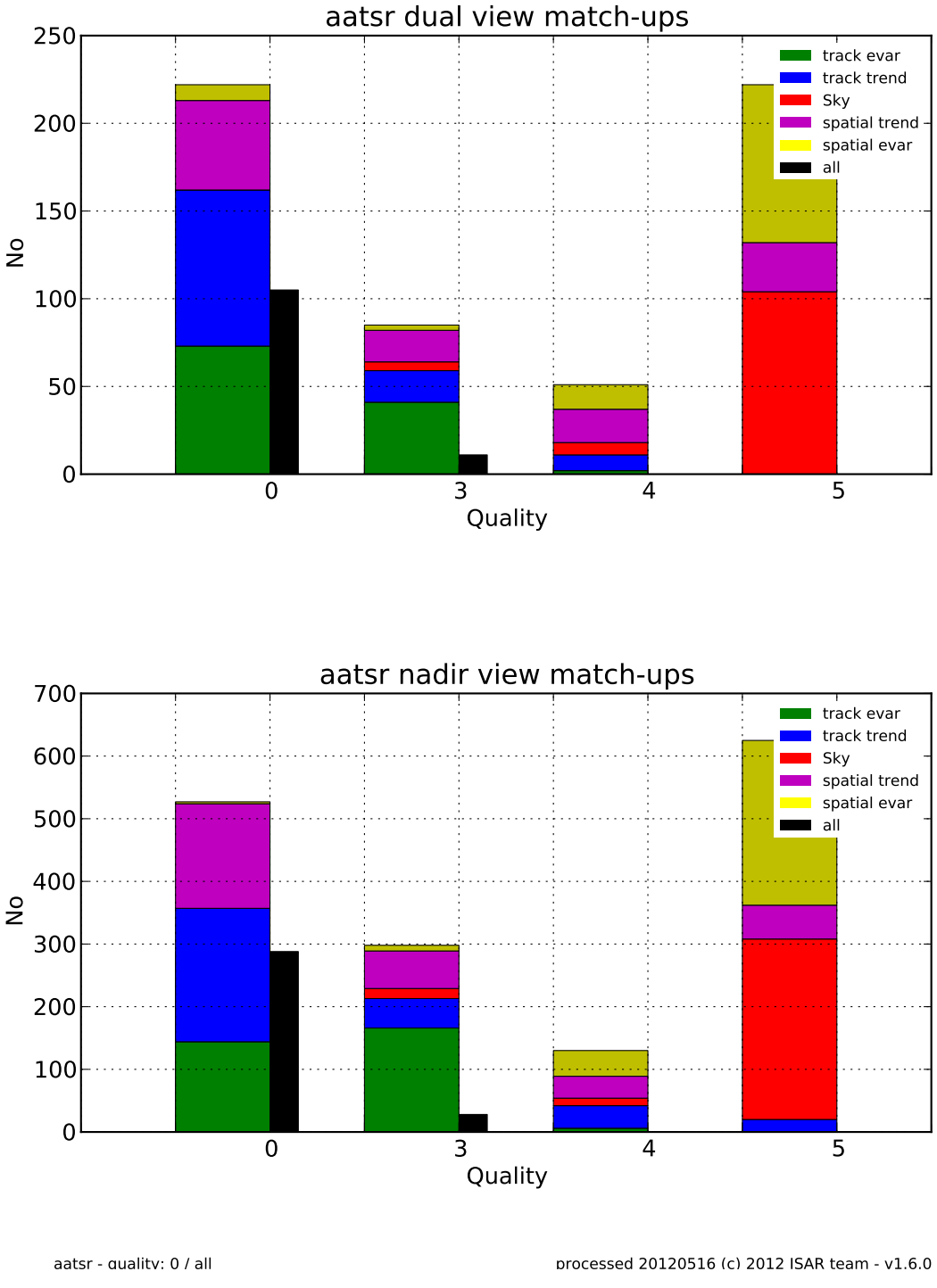


FIGURE 6.8: Quality indicator plot showing how the different parameters set the quality level for the *match − up* pair for the AATSR day time *match − up* from the training data set. Note that the bar plots were stacked on top of each other for better visibility and only show which indicator has the most impact in the given scene.

The thresholds used for the processing of the whole data set are shown in table 6.2.

TABLE 6.2: Threshold values of indicators, I_x, for assigning quality flags Q_x

Quality flag	Units	0	3	4	5
Q_{P1}	K^2	No threshold	0.3	0.1	0.035
Q_{P2}	K^2	No threshold	0.3	0.1	0.035
Q_T	K	No threshold	0.6	0.2	0.05
Q_S	K	No threshold	0.25	0.07	0.025
Q_{sky}	K	No threshold	280	260	240

For Q_X to be assigned the value at the top of the column, the estimate of I_X must be less than the number given in the row for Q_X. Thus, in practice, each indicator , X, is tested by entering the table from the right in row Q_X. If the estimate of I_X is smaller than the threshold given, then Q_X is assigned the best quality value of 5. If it is greater than the given threshold, the test is applied in the column to the left and so on until the $Q_X = 0$ column is reached. This is repeated for each indicator until each $match - up$ point has a set of values between 0 and 5 for each of Q_{P1}, Q_{P2}, Q_S and Q_T. To determine the overall quality value, Q, for each match-up pair, the lowest value of the set of Q_X is selected, that is:

$$Q = min(Q_{P1}, Q_{P2}, Q_S, Q_T, Q_{sky}) \tag{6.9}$$

In other words the overall QI value is set by the greatest of the predicted sources of uncertainty for that particular $match - up$ pair. It must be noted that the indicator I_X, as estimated by equations 6.5 to 6.8 are estimates of the magnitude of the corresponding uncertainties. They are not an actual estimation of the uncertainty for each of the parameters discussed in equation 6.4. However the main aim of this approach is to design a QI which helps to stratify the $match - up$ dataset and thus enables the user of such a data set to use only those data with the highest quality. Furthermore it must be noted that the thresholds, while estimated from real data with an empirical method, can not be directly mapped to an interpretation of the underlying oceanographic processes at the $match - up$ scene. The thresholds were mainly designed to reduce overall uncertainty in the $match - up$ process and are in some cases overzealous as can bee seen in figures 6.5 and 6.7. Ideally the standard deviation will become lower the higher the QI value, which

is demonstrated for the data set discussed in section 6.6. Note that as a starting point for the QI data processor, grade 4 data (in terms of the classical space time coincidence windows - see table 6.1) was used which means the data going into the QI process was already limited to be within ± 6 h and ± 25 km of each other. Therefore a $Q = 0$ in table 6.4 is the same as a grade 4 in table 6.1. Please also note that the populations for each QI level above 0 are exclusive, in other words QI 3, 4 and 5 have only data from the respective level and do not include data from the higher level.

6.6 Results

The results of applying the QI method to the *match−up* data between AATSR and ISAR are stratified by statistical method and GHRSST CV value. Two statistical methods were used, normal Gaussian statistics and robust statistics (Daszykowski *et al.*, 2007). The robust statistics were used to see the effect of using a mathematically more elegant method of treating outliers and results are shown with a Huber T distribution. The median absolute deviation (MAD) as calculated by the robust statistics has been converted by using equation 6.10 into a robust standard deviation (RSD) equivalent for easier comparison with the normal statistics results.

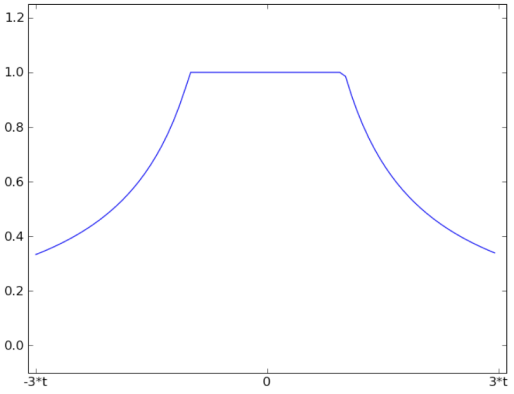


FIGURE 6.9: Robust Huber T distribution.

$$\begin{aligned} \sigma &= K \cdot MAD = \text{RSD} \\ K &= \frac{1}{\phi^{-1}\left(\frac{3}{4}\right)} \approx 1.428 \end{aligned} \tag{6.10}$$

where σ represents the standard deviation, K is a constant scale factor and ϕ^{-1} the normal inverse cumulative distribution function.

TABLE 6.3: Quality indicator results for 2004 to 2009 summary table. Table labels starting with B can be found in appendix B and table labels starting with a 6 can be found in this section.

Data CV	Statistics	Table
0	Normal	B.2
	Robust	B.6
3	Normal	B.3
	Robust	B.7
4	Normal	B.4
	Robust	B.8
5	Normal	B.5
	Robust	6.4

A full set of validation results for all confidence values of AATSR have been evaluated using the new $match - up$ quality indicators. Only results from validating the AATSR CV 5 data with Robust statistics are discussed here. The remaining results can be found in the appendix, using table 6.3 as a key to where they are located in Appendix B. Each results section has first a table with the $match - up$ statistics (tables 6.4, B.2, B.6, B.3, B.7, B.4, B.8 and B.5) followed by a number of plots showing the $match - up$ data in graphical form.

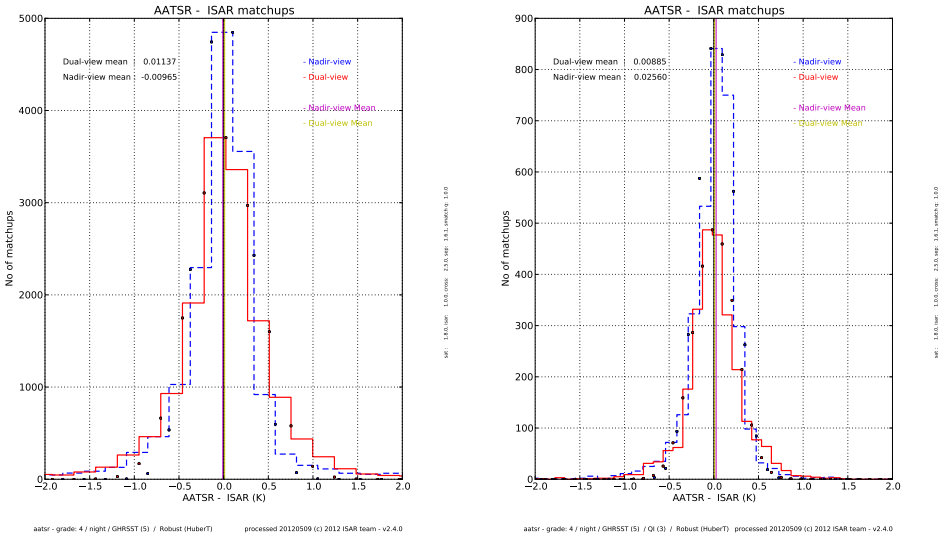
The data in table 6.4 shows the data stratified by QI for the AATSR dual view and nadir view. Both the AATSR dual view and nadir view section first show the results for both the two channel and three channel data combined and than expanded into the two channel data and three channel data results. In table 6.4 the first column shows the quality level, the second column the mean difference between the $match - up$ pairs (the mean value is represented by the median estimation in the robust statistics case.), third column the standard deviation (or RSD in the robust statistics case), fourth column the number of $match - up$ pairs, the fifth column the number of satellite overpasses providing the pairs, the sixth and the seventh column the minimum temperature and maximum of temperature measured by a $match - up$, the eight and ninth column the minimum and maximum time difference between a $match - up$ pair (Note the time difference is not signed and therefore does not indicate whether the AATSR or ISAR was measured

earlier), and the tenth and eleventh column show the minimum and maximum distance between the AATSR and ISAR *match − up* .

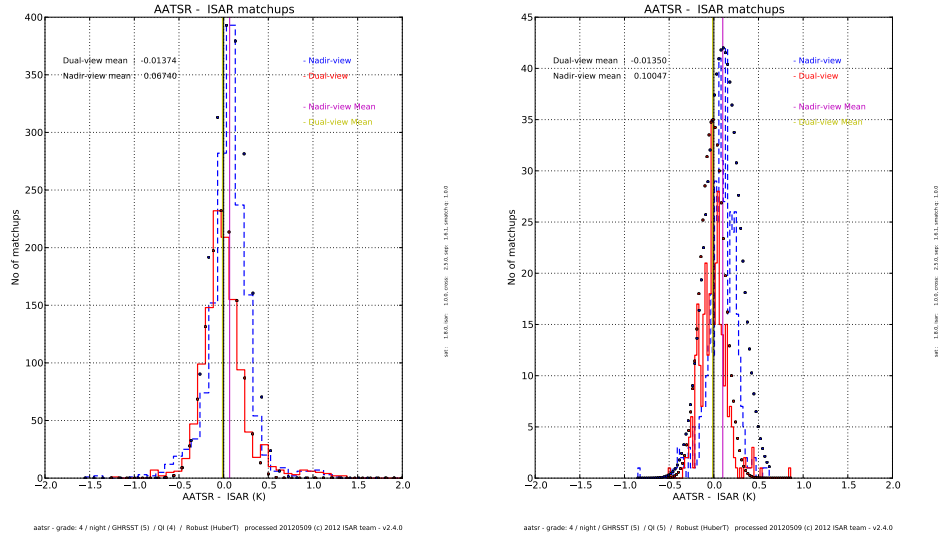
The plots in figures 6.10, 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, 6.17, 6.18, are grouped in such a way that each group shows the plots for QI 0 (a), QI 3 (b), QI 4 (c) and QI 5 (d). Furthermore the analysis results are presented separately for day (2 channel) and night (3 channel) data and for dual view and nadir data. In the plots section first the histograms of the data are shown split in night time, shown in figure 6.10 and day time data, shown in figure 6.11. In both figures dual view data is coloured red and nadir view data is coloured blue. The second set of plots are the scatter plots split into day, shown in figure 6.13 and night time data, shown in figure 6.12 for the dual view AATSR data. The nadir view day time scatter plots are shown in figure 6.15 and the night time nadir view scatter plots are shown in figure 6.14. The third set of plots are the location plots indicating where the *match − up* has occurred as shown in figure 6.16. The fourth set of plots are the distance and time histograms showing how far apart in the spatial and temporal dimension the data was matched. Figure 6.17 shows the night time and figure 6.18 the day time results, with dual view *match − up* pairs shown in red and nadir view *match − up* pairs shown in blue. Note the very narrow distance histogram is a function of the *match − up* process where first distance and then time is minimised.

TABLE 6.4: Quality indicator results for 2004 to 2009 for AATSR CV 5 data and robust statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	0.04	0.44	24705	442	273.92	298.39	0.02	359.98	0	0.32
q3	0.02	0.27	3796	252	273.92	298.39	0.17	359.80	0	0.32
q4	0.00	0.21	1631	156	273.92	298.39	0.23	359.85	0	0.28
q5	0.00	0.15	425	75	273.92	298.39	0.05	355.37	0	0.30
2 channel view										
q0	0.09	0.55	9815	208	273.92	297.56	0.02	359.98	0	0.32
q3	0.05	0.36	1233	104	273.92	297.56	0.63	358.90	0	0.32
q4	0.04	0.31	469	58	273.92	297.56	0.27	359.50	0	0.28
q5	0.03	0.25	110	29	273.92	297.56	0.55	317.03	0	0.30
3 channel view										
q0	0.01	0.39	14890	234	277.11	298.39	0.02	359.90	0	0.32
q3	0.01	0.24	2563	148	277.11	298.39	0.17	359.80	0	0.32
q4	-0.01	0.18	1162	98	277.11	298.39	0.23	359.85	0	0.24
q5	-0.01	0.13	315	46	277.11	298.39	0.05	355.37	0	0.22
Nadir view										
q0	0.27	0.50	24705	442	273.92	298.39	0.02	359.98	0	0.32
q3	0.22	0.31	4954	252	273.92	298.39	0.13	359.80	0	0.32
q4	0.22	0.30	2169	156	273.92	298.39	0.18	359.85	0	0.31
q5	0.24	0.30	630	75	273.92	298.39	0.02	358.98	0	0.30
2 channel view										
q0	0.78	1.22	9815	208	273.92	297.56	0.02	359.98	0	0.32
q3	0.71	1.10	1708	104	273.92	297.56	0.18	358.90	0	0.32
q4	0.67	1.07	651	58	273.92	297.56	0.27	359.50	0	0.29
q5	0.68	1.04	196	29	273.92	297.56	0.55	358.98	0	0.21
3 channel view										
q0	-0.01	0.28	14890	234	277.11	298.39	0.02	359.90	0	0.32
q3	0.03	0.21	3246	148	277.11	298.39	0.13	359.80	0	0.32
q4	0.07	0.19	1518	98	277.11	298.39	0.18	359.85	0	0.31
q5	0.10	0.19	434	46	277.11	298.39	0.02	355.37	0	0.30



(a) (b)



(c) (d)

FIGURE 6.10: QI histograms for CV 5 and robust statistics for night time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view *match − up* data, the dashed blue histogram shows nadir view only *match − up* data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

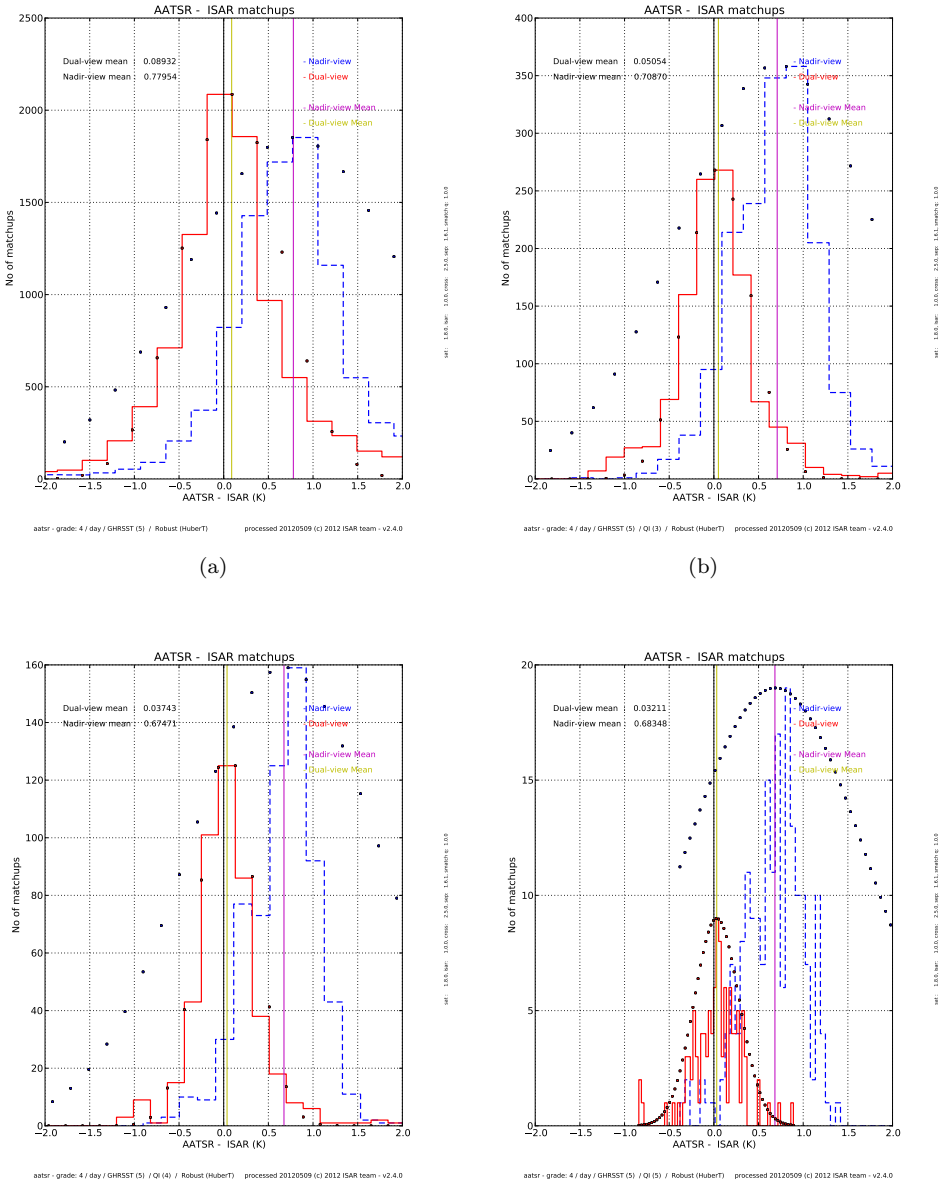


FIGURE 6.11: QI histograms for CV 5 and robust statistics for day time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view *match − up* data, the dashed blue histogram shows nadir view only *match − up* data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

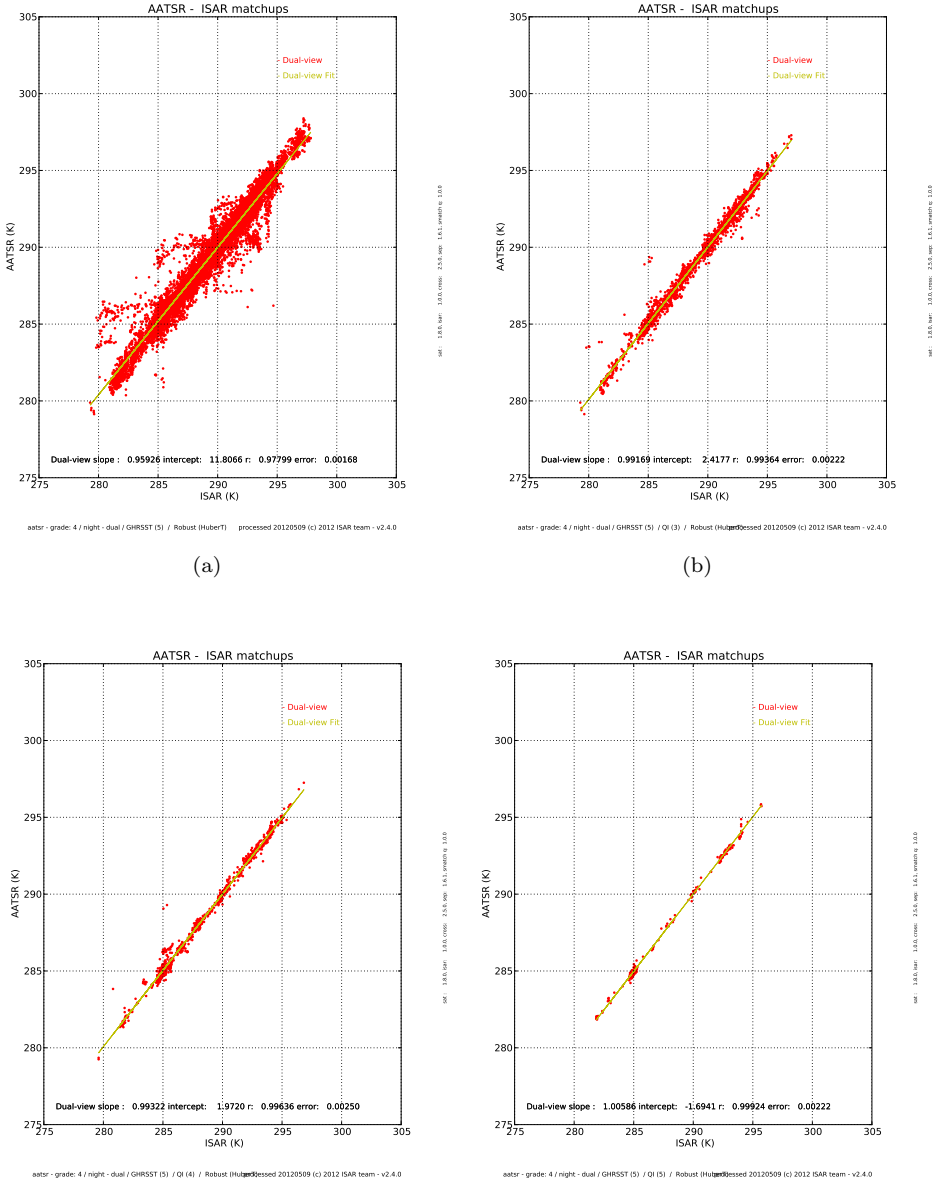


FIGURE 6.12: QI scatter plot for CV 5 and robust statistics for dual view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

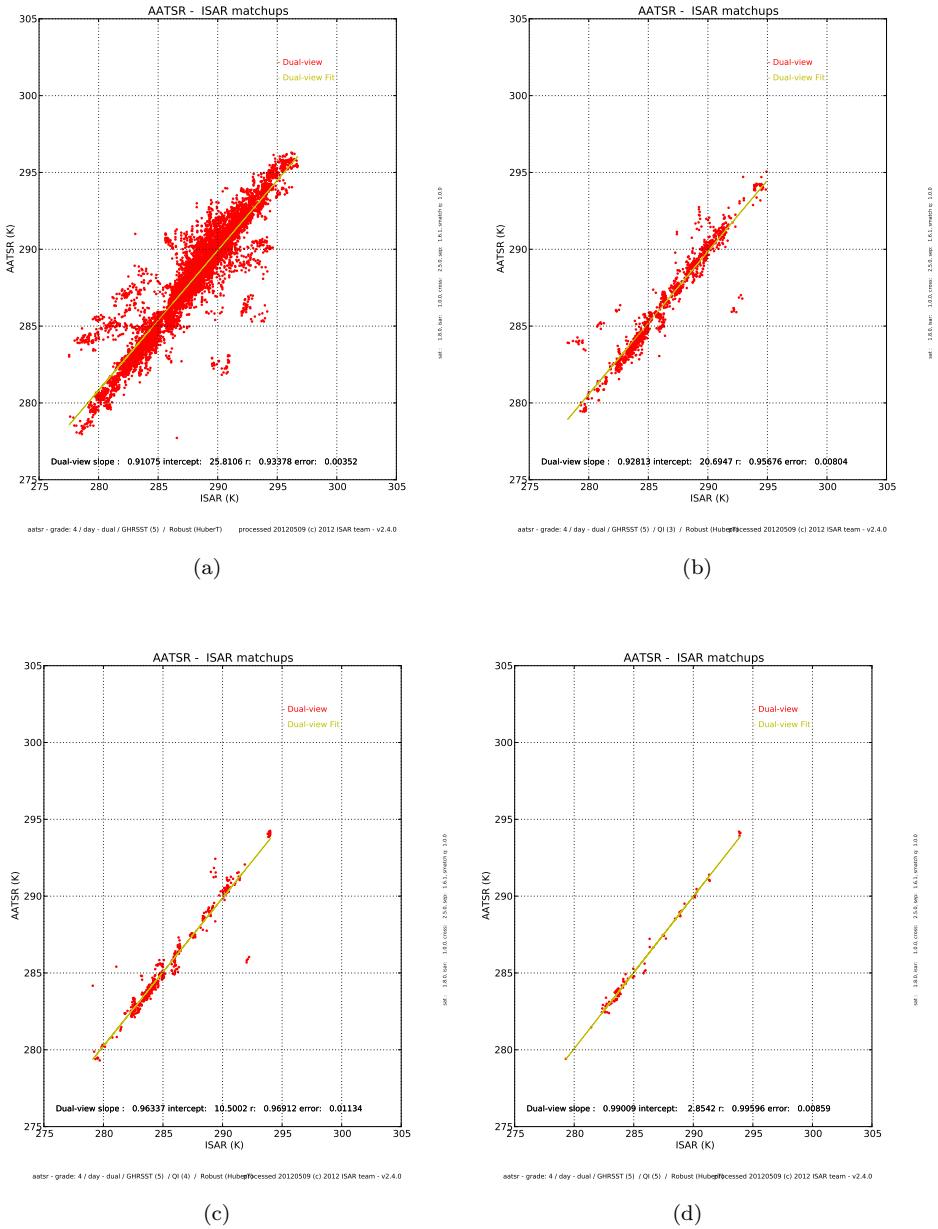


FIGURE 6.13: QI scatter plot for CV 5 and robust statistics for dual view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

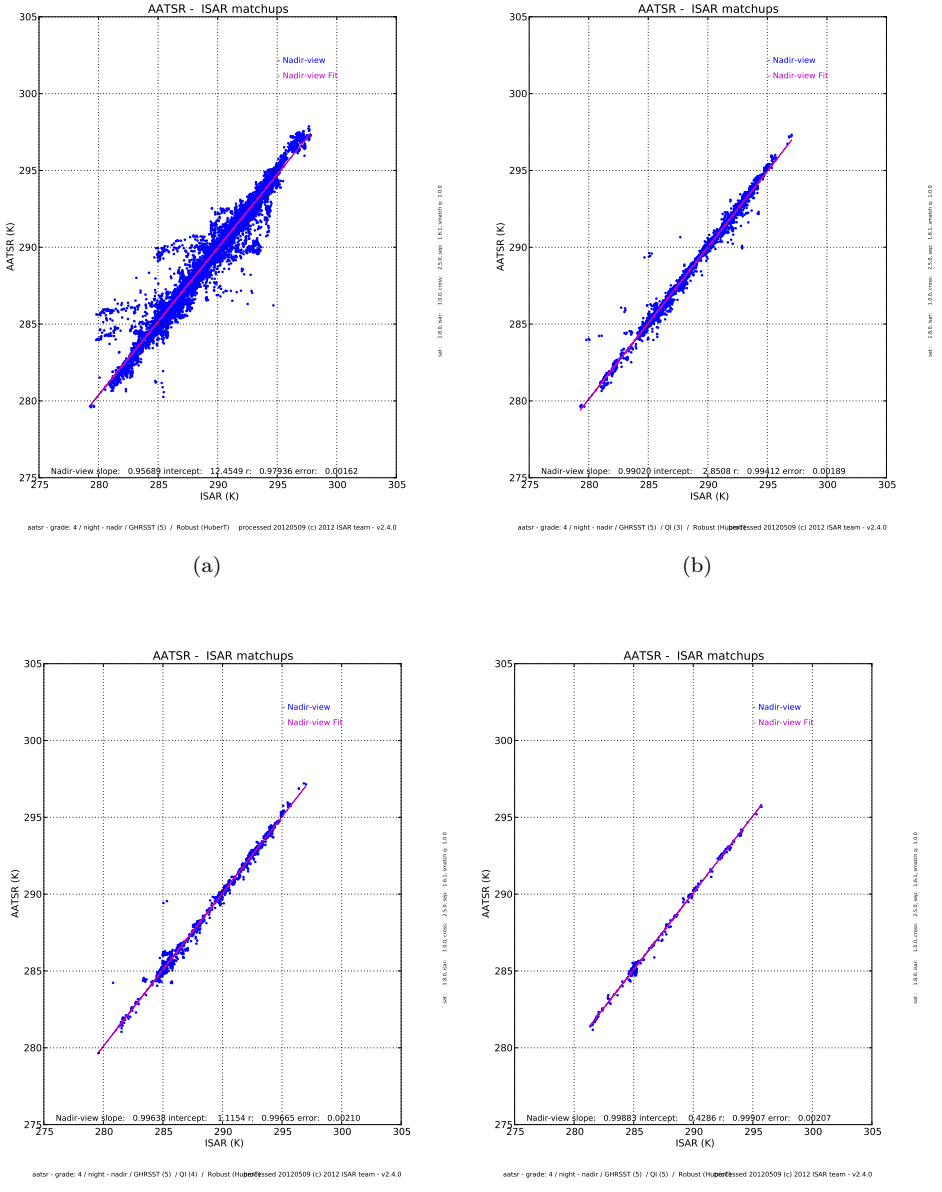
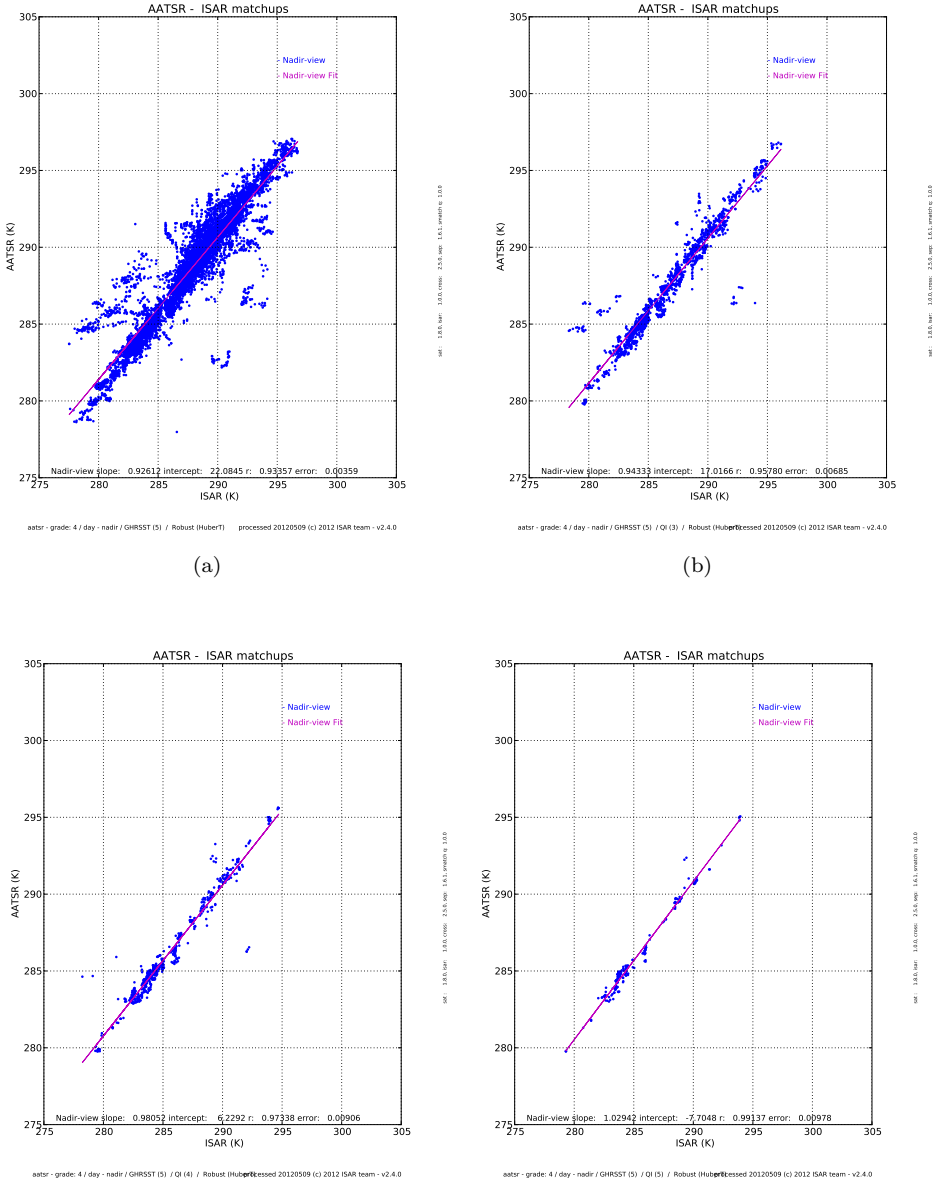


FIGURE 6.14: QI scatter plot for CV 5 and robust statistics for nadir view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;



(a) (b)

(c) (d)

FIGURE 6.15: QI scatter plot for CV 5 and robust statistics for nadir view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

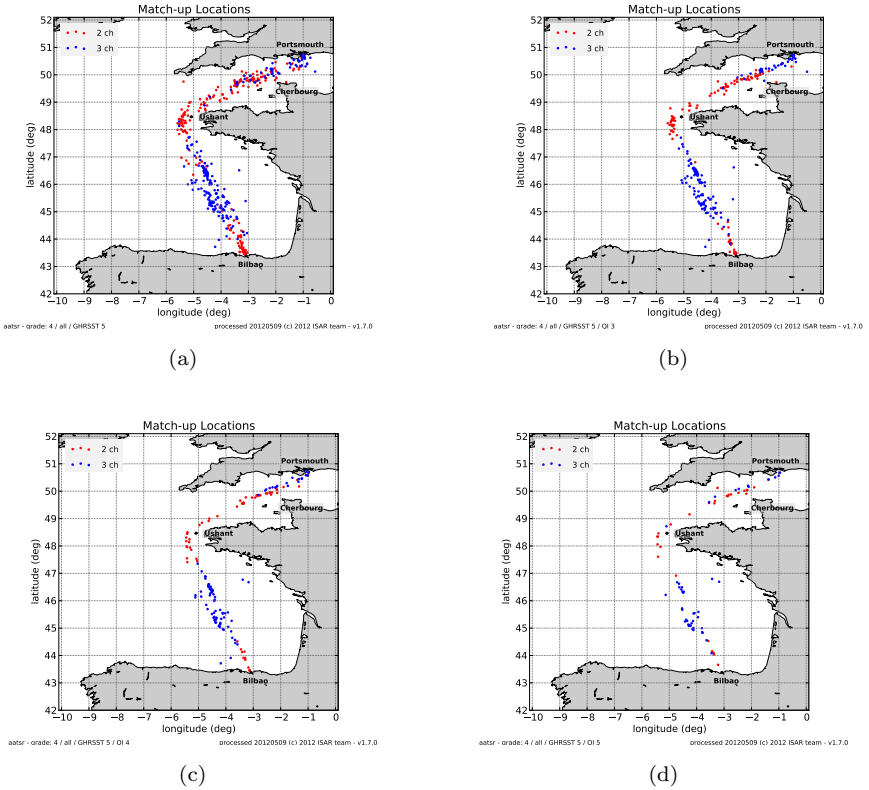
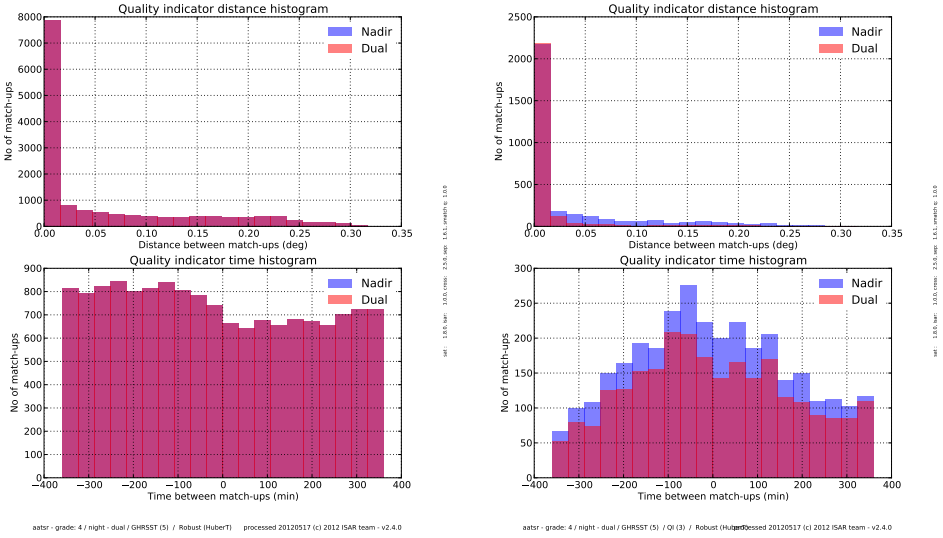
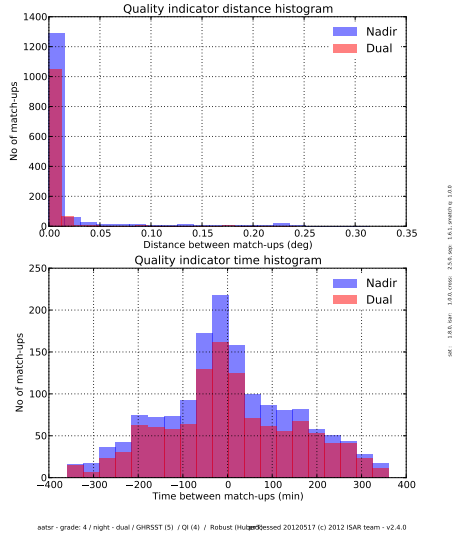


FIGURE 6.16: QI location plot for CV 5 and robust statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; (c) QI 4 location plot; and, (d) QI 5 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue

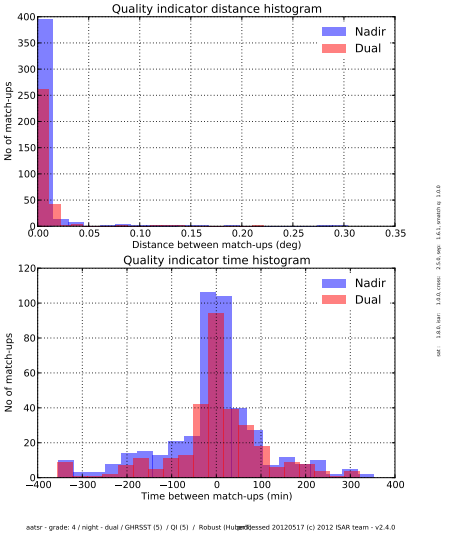


(a)

(b)

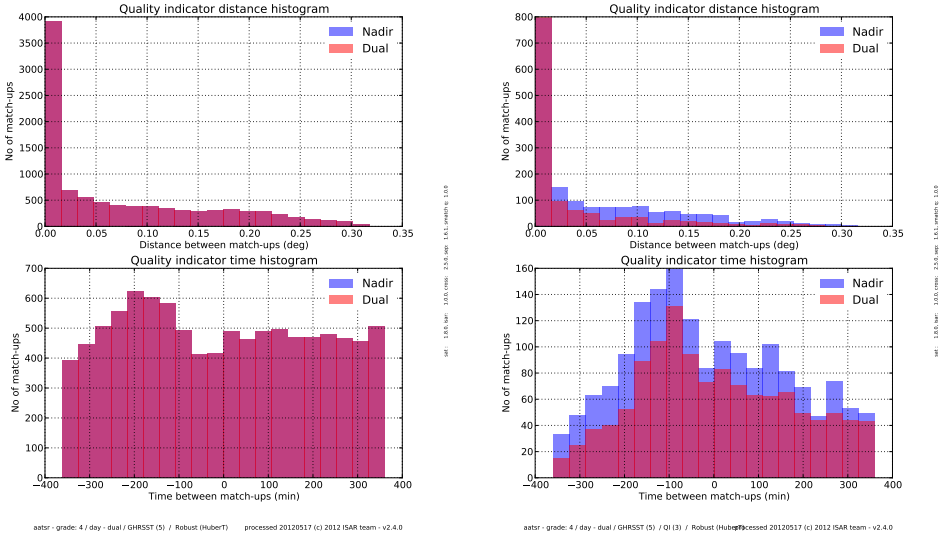


(c)



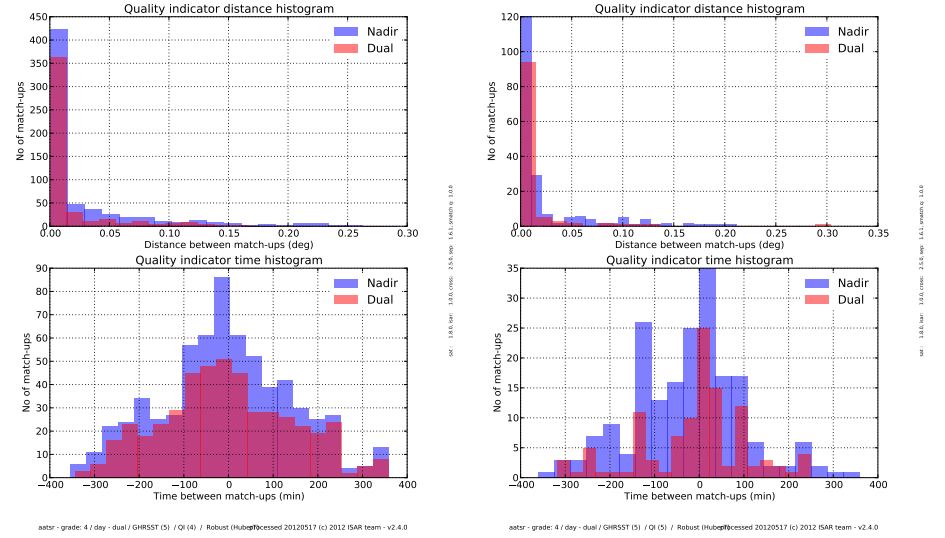
(d)

FIGURE 6.17: QI time and distance plot for CV 5 and robust statistics for for night time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match-up$ data are coloured red and nadir view time $match-up$ data are coloured blue



(a)

(b)



(c)

(d)

FIGURE 6.18: QI time and distance plot for CV 5 and robust statistics for for day time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue

6.7 Discussion of the results

The results in table 6.4 show that the QI works in most cases as desired, by reducing the uncertainty as the QI value increases. However in some cases, like the nadir view 2 channel and 3 channel data the step from QI 4 to QI 5 is not improving the standard deviation and slightly increasing the bias. This phenomenon can also be seen in tables B.6, B.7 and 6.4, which seems to be partly related to using robust statistics. The same tables for normal statistics (table B.2, B.3, B.4 and B.5) show the slight increase in bias but a decrease in standard deviation. The increase in bias is mainly due to the balancing effect some of the outliers seem to have, which can be observed in the scatter plots (e.g.: figure 6.12, outliers either side of the 1:1 relationship result in a near zero bias.). The scatter plots in figures 6.12, 6.13, B.44, B.31, B.32 also show that the width of the scatter plot around the 1:1 line gets reduced by the increasing QI value.

Figures 6.16, B.7, B.13, B.18, B.26, B.35, B.41, B.46 show a good distribution of the $match - up$ location throughout the ship track. The stratification of day time and night time $match - up$ location is a side effect of the PoB schedule. The time and distance plots (figures 6.17, 6.18, B.36, B.37, B.42, B.47, B.8, B.37, B.14, B.19, B.27, B.28) show that the $match - up$ data in most cases is within the same pixel or very close by. The temporal data $match - up$ on the other hand shows that data across the whole temporal window is used in the QI 0 case with a small bias towards $match - up$ pairs where the ISAR measurement leads the AATSR measurement. This small negative bias changes to a zero time difference in the QI 5 cases.

Overall the QI method does quantify the uncertainty of the $match - up$ process and produces good results for all CV 5 and CV 0 cases. There are some issues with the method for the CV 3 and 4 $match - up$ data, which is mainly due to the small number of $match - up$ pairs in those two cases of 260 and 750 respectively. Also while the method tries to reduce the $match - up$ uncertainty and therefore $match - up$ outliers, it still classifies a number of $match - up$ pairs where the difference between AATSR and ISAR derived temperature is very small as a low quality $match - up$. This is due to the fact the indicators as described in section 6.5 determine situations where the uncertainty of

the $match-up$ is high, they do not give any information about the actual difference between the AATSR and ISAR derived temperature.

The main reason for still having some $match-up$ outliers in the QI stratified data is due to cloud contamination in the AATSR data (Hopkins, 2010) and that the QI method has no direct parameter for the AATSR cloud probability. The ISAR BT_{sky} measurement, while improving the situation, has the same issue as the SST $match-up$ data that it is not measured at the same location and time as the AATSR acquires its data. Figure 6.19 shows an example of where the QI method did not catch all of the questionable $match-up$ pairs. The $match-up$ processor found 144 night time dual view $match-up$ pairs, with 3 being classified as QI 4, 21 as QI 3 and all the other $match-up$ pairs as QI 0. Particularly two of the $match-up$ pairs that were classified as 3 stand out with a difference between the AATSR dual view SST_{skin} and the ISAR SST_{skin} of -0.75 K. The cold AATSR points towards a cloud contamination issue not picked up by the method. The ISAR BT_{sky} shows in both cases clear sky conditions for the ISAR measurement.

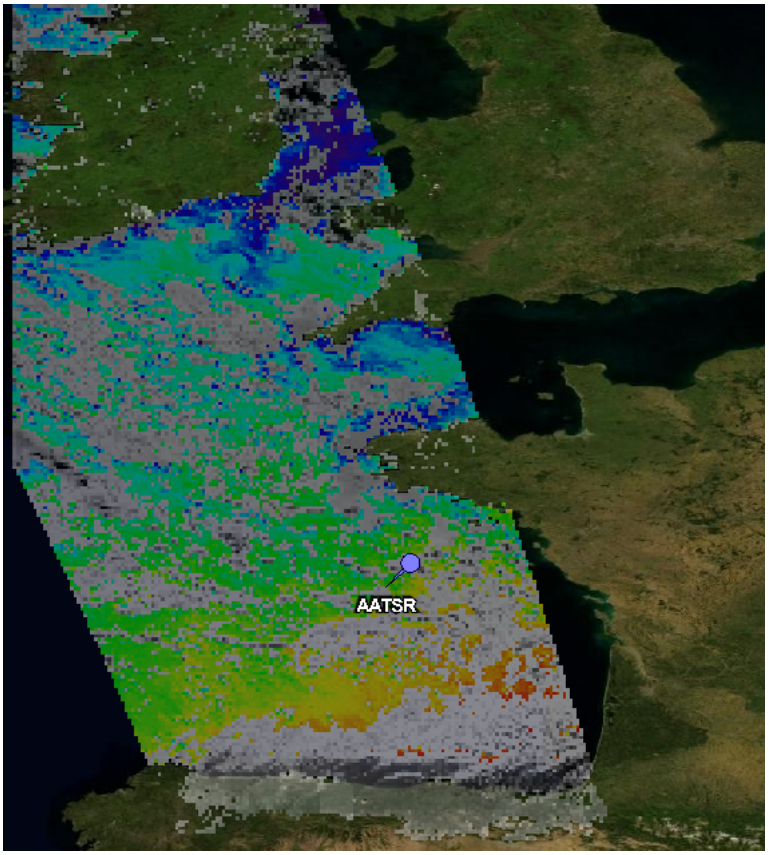


FIGURE 6.19: AATSR image from 24.09.2009, showing a $match-up$ scene with cloud contamination issues.. From Hopkins (2010).

6.8 Implications of the QI

In this chapter a method for estimating the $match - up$ uncertainty was presented. The uncertainty is estimated via a set of QIs that describe the sea surface variability at the acquisition time of the two measurements that are used for the comparisons. The QI resolve the variability well and show for most cases the desired effect of reducing the uncertainty with higher QI value. While there are some issues, such as reducing also good data points in the quest to reduce the $match - up$ outliers it compares well with a classic $match - up$ window validation method. A comparison with Wimmer *et al.* (2012) shows that QI 3 compares well with grade 2b (± 2 h and ± 0.01 deg) with a comparable standard deviation but higher number of $match - up$ data points. Similar, QI 4 compares quite well with grade 1 $match - up$ data with similar standard deviations and $match - up$ numbers. Tables 6.5 and 6.6 show the comparison of the two methods for the two mentioned cases.

TABLE 6.5: Comparison between QI 3 results and Wimmer *et al.* (2012) grade 2b results for CV 5 and robust statistics.

AATSR Data	QI 3			Grade 2b		
	mean	std	No	mean	std	No
D3	0.01	0.24	2563	0.08	0.22	2462
D2	0.05	0.36	1233	-0.01	0.39	1024
N3	0.03	0.21	3246	0.08	0.21	2462
N2	0.71	1.10	1708	0.65	0.99	1024

TABLE 6.6: Comparison between QI 4 results and Wimmer *et al.* (2012) grade 1 results for CV 5 and robust statistics.

AATSR Data	QI 4			Grade 1		
	mean	std	No	mean	std	No
D3	-0.01	0.18	1162	0.03	0.19	707
D2	0.04	0.31	469	-0.01	0.33	238
N3	0.07	0.19	1518	0.08	0.19	70
N2	0.67	1.07	651	0.64	0.99	238

From the comparison with results of Wimmer *et al.* (2012) it could be argued that QI 5 takes the reduction in $match - up$ uncertainty a bit too far. However as the QI levels were not designed to match classic validation limits but to quantify the $match - up$ process uncertainty and to that extent examine how low the $match - up$ uncertainty can be,

the limits for QI 5 and subsequently all the other levels were not changed. The results also show that the *match − up* uncertainty for QI 5 data is well below the AATSR design accuracy of 0.3 K. In fact QI 5 results go some way to validate the claim that AATSR data is of quality required for climate studies and should meet the requirement of absolute uncertainty of 0.1 K and a stability of 0.04 K $decade^{-1}$ (Ohring *et al.*, 2005).

Chapter 7

Conclusion

In this section first a summary of the science chapters is presented, then the scientific questions presented in chapter 1 are revisited, and finally a discussion of potential future work is given.

7.1 Summary and Discussion of the main achievements

7.1.1 ISAR uncertainty

To investigate the impact of the *in situ* SST_{skin} measurements, an uncertainty model for the ISAR instrument was developed as reported in chapter 4. This is the first time this has been done for a seagoing infra-red radiometer from first principles, assigning an uncertainty to each element of the measurement process, not just for the instrument as a whole. First the main components of the ISAR instrument, which contribute to the uncertainty the SST_{skin} measurement were identified. Then each instrument component was assigned an uncertainty which was propagated through to the final SST_{skin} value. This uncertainty model enables an uncertainty value to be estimated for each SST_{skin} measurement. which previously was not possible. The main contributions to the instrument uncertainty are the BB thermistors and the emissivity of the sea water. While the uncertainty of the thermistors in the BBs defines the base line uncertainty

of the instrument, it is very well defined. On the other hand the sea water emissivity uncertainty is more difficult to estimate as it depends not only on the emissivity model, but also on the sea state and viewing angle. The results shown in figures 4.8 and 4.9 indicate that the estimated ISAR uncertainties are mainly below the quoted instrument uncertainty of $\pm 0.1\ K$ (Donlon *et al.*, 2008). The main reason for the occasional high uncertainties ($> 0.1K$) is a variable target view detector signal. This mainly occurs in ports (see figure 4.9, panel c) or areas of quickly changing *SST* due to shipping or moving fronts. It must also be noted that, presented in this form, the ISAR uncertainty is a measurement uncertainty and not just an instrument uncertainty. The use of this more rigorous uncertainty model means that the quality of each individual ISAR record is assessed. Depending on the particular application of their use, for example for satellite validation with a given precision requirement, records can be discarded if their quality does not meet the requirement. Following this work, it is intended that in future the estimation of ISAR measurement uncertainties should become an integral part of the ISAR *SST* retrieval software. The standard, GHRSST endorsed, international format for archiving *SST* measured by shipborne radiometers will in future include a field for recording the uncertainty of each individual record.

7.1.2 Uncertainty analysis of *SST* measurements

In chapter 5 the uncertainty of different measurements were estimated by the means of statistical estimation. Three different methods were compared, with Tokmakian and Challenor (1999) giving the most consistent results which is due to the combination of bias and random component. The uncertainty estimates for AATSR, ISAR and SBE 48 show that ISAR has the lowest uncertainty ranging from 0.09 K to 0.14 K, where the Challenor (2004) method gives a more stable range from 0.09 K to 0.1 K. This is consistent with the typical magnitude of uncertainties calculated directly for each ISAR record following the method developed in chapter 4. AATSR uncertainties range from 0.04 to 0.2 K with again Challenor (2004) showing the lower results. The AATSR uncertainty estimation also showed that the N3 results have the lowest uncertainty and the N2 the highest uncertainty with D3 and D2 uncertainties in the middle of the range. These

results are comparable with O'Carroll *et al.* (2008), Wimmer *et al.* (2012) and Corlett *et al.* (2006). The SBE 48 uncertainty estimates have a range from 0.19 to 0.23 K, which is comparable to drifting buoys (Beggs *et al.*, 2012). This confirmation that the ISAR measures SST_{skin} with less uncertainty than hull thermometry also demonstrates that the use of shipborne radiometers is critical for satellite SST validation, as demonstrated by Wimmer *et al.* (2012) and Corlett (2012).

7.1.3 Uncertainty analysis of the $match-up$ process

The uncertainty of the validation process itself was investigated in chapter 6. The uncertainties of the $match-up$ process were analysed, which led to the definition of a set of indicators describing the quality of each $match-up$ pair. The overall $match-up$ quality indicator, QI, is a quality classification assigned to each $match-up$ pair individually using a set of objective scientific criteria based on oceanographic and atmospheric conditions and which are mainly independent of the actual $match-up$ comparisons. This is in contrast to the classic validation method in which all data within a predefined $match-up$ window is presumed to be of equal quality and used for filtering the computation of the $match-up$ statistics. The validation results of AATSR and ISAR data from 2004 to 2009 were analysed with two different statistical methods as well as stratified by the AATSR CV level. The QI 3 results showed a similar mean and standard deviation in comparison to a grade 2b in Wimmer *et al.* (2012) and the QI 4 showed a comparable mean and standard deviation in comparison to a grade 1 in Wimmer *et al.* (2012), while both QI 3 and QI 4 have a higher number of $match-up$ pairs than the grade 2b and grade one results respectively in Wimmer *et al.* (2012).

The comparison with Wimmer *et al.* (2012) shows clearly that the new QI method not only yields higher $match-up$ numbers but also produces improved statistics by identifying and excluding $match-up$ pairs where oceanographic or atmospheric conditions are not ideal for comparisons. Therefore using the QI method to populate $match-up$ databases will improve the validation of satellite sensors compared to the $match-up$ window method as the uncertainty of the $match-up$ process can be estimated and as a consequence not added to satellite uncertainty budget.

7.2 Scientific relevance

In section 1.4 the underlying scientific questions motivating this thesis are stated to be:

Q.1 How do conventionally measured SST and its uncertainties impact the climate record? For this study the "conventionally measured" SST refers to shipborne SST_{depth} measurements.

Q.2 Can SST_{skin} measurements acquired by satellites be a better data source for a climate record than buoy measured SST_{depth} ?

Q.3 To what extent has a poor choice of validation data and validation method led to under-appreciation of the quality of satellite acquired SST_{skin}.

To address the scientific questions three specific aims of this thesis were identified as:

A.1 Assessing the uncertainties associated with SST_{skin} measurements systems, using ISAR as an example.

A.2 Evaluating the uncertainties related to different measurement techniques on board VOS and SOO, especially in relation to SST_{depth} and SST_{skin} measurements. This includes the use of advanced uncertainty calculation methods developed by, for example, Tokmakian and Challenor (1999) and O'Carroll *et al.* (2008).

A.3 Examining the uncertainties introduced by validation methods, which permit some latitude in the specification of the spatial and temporal coincidence of samples from different datasets being inter-compared. The focus here is on the uncertainty budget of the $match - up$ process.

The specific aims were designed to cast light on some of the issues underlying the scientific questions.

The third scientific question can be quite clearly answered by this study. A poor choice of validation data can lead to the mistaken conclusion that the satellite derived SST is

of poorer quality than it actually is. This is particularly likely with a classic validation approach in which it is impossible to separate the uncertainties associated with the satellite measurement from the uncertainties of the *in situ* measurement and the *match−up* process. The methods developed in chapter 4 and 6 can now be used to estimate the uncertainties associated with each measurement and to some degree uncertainties associated with the method of matching pairs of data for validation which, if ignored can introduce an additional uncertainty of up to 0.2 K (Minnett, 1991). The validation results for AATSR shown in chapter 6, when compared with those already published by Wimmer *et al.* (2012), show reduced differences between the satellite and the validation data, at least within the geographical limits of this study, which should enhance overall confidence in the reliability of the AATSR global dataset for use in operational and climate applications.

For the second scientific question the answer is not quite as clear as for the third question. Projects like ARC (Merchant *et al.*, 2008) have shown that AATSR data can be considered as a CDR (Embury *et al.*, 2012a). Satellite derived *SST* is more consistent, than *SST* based on conventional ship and buoy observations, especially if there are overlap periods between different generations of the same sensor. However the satellite *SST* data record is still relatively short, spanning no longer than 30 years, and for climate application it will have to be used together with the *in situ* data record for *SST* which is available for longer time scales. Minnett and Corlett (2012) state that in order for satellite derived *SST* to be considered as a CDR not only the validation of the *SST* measurement is important but also the validation of the satellite sensor uncertainty model. At present this is only achievable by using traceable *SST* measurement as recorded by ISAR. This is where the per SST_{skin} measurement uncertainty calculation of the ISAR system is very valuable and will help to demonstrate the validity of satellite *SST* as a CDR.

The first scientific question is difficult to answer directly with this study. However in chapter 5 it was shown that both the satellite and *in situ* measured SST_{skin} have lower uncertainties than the SST_{depth} measured by a contact thermometer. Studies by Kennedy *et al.* (2011); Kent and Berry (2005); Kent and Challenor (2006); Rayner *et al.*

(2006), to name a few, show that conventionally measured *SST* can be fit for purpose as a CDR. In order to achieve the CDR requirement of 0.1 K accuracy and a stability of 0.04 $K\ decade^{-1}$ (Ohring *et al.*, 2005) a number of corrections for systematic and random uncertainties have to be made. These corrections are mainly needed because of the missing traceability of either the temperature probe or the temperature measurement system or both. This is especially true for historic measurements and, while modern temperature probes are calibrated before use, very few are calibrated after a deployment, especially when the probe is used on buoys. While a clear assessment of this question is outside the scope of this thesis, it seems from the work presented here that, at least on SST_{skin} measurements of a CDR, radiometric measurements by *in situ* or satellite sensors have a lower uncertainty than those measured by conventional means. The long term stability of radiometric measurement is more difficult to assess than the accuracy of the measurement, because of the fairly short record lengths of AATSR and ISAR data. However from the traceable ISAR calibration records it has been shown that the AATSR is stable over the data record period of eight years. Also the inter-comparison experiments in 2001 and 2009 (Barton *et al.*, 2004; Rice *et al.*, 2004; Theocharous and Fox, 2010; Theocharous *et al.*, 2010) have shown no drift in the ISAR measurements or the calibration target CASOTS II.

7.3 Future work

The work shown in this study demonstrates how the uncertainties associated with SST_{skin} measurements can be assessed. Not only was a method to assess measurement uncertainties developed but also a method to estimate the uncertainty introduced by the validation method.

The research work begun in this thesis needs to be extended in two ways. First there are aspects of the new methodology that can be refined and possibly improved. Second it is important that the new methods for creating *in situ match − up* databases for satellite *SST* products with quantified uncertainty be applied to existing and future *SST* products to achieve more reliable validations of those products.

ISAR uncertainty The instrument uncertainty model presented in this thesis is a first approach to produce an uncertainty model which produces an uncertainty for each measurement recorded by ISAR by looking at all the contributing factors. While the uncertainty model works well, there are a few issues still worth investigating to improve the uncertainty estimation. The first issue with the current approach is that there are a few unresolved correlation issues between Type A and Type B as well as between measurement and instrument uncertainties. This is mainly due to the fact that the uncertainty model relies a great deal on information from the literature and the only way to address some of the correlation issues would be to replicate some of the work in the literature with dedicated experiments, which was not possible for this thesis. Furthermore at present the mirror gain factor, as discussed in section 3.2.4, is not considered in the uncertainty model and should be included in a future version. While pre- and post-deployment characterisations of the ISAR instrument are also not considered in the uncertainty model, they can instead be used to verify the uncertainty model.

$Match - up$ **quality indicator** The QI depend on a fixed set of thresholds which have been derived using a training dataset of coincident AATSR data. To make wider use of the QI method, especially using it on other satellite SST, the thresholds need to be either derived for each sensor or, more elegantly, an objective method could be designed to estimate the thresholds automatically. Furthermore the ISAR instrument uncertainty should be incorporated into the QI method to help estimate the AATSR uncertainty.

The ISAR uncertainty for each SST_{skin} measurement together with the QI method are currently not part of the standard AATSR $match - up$ processor. To take full advantage of both new methods first the full ISAR data record from 2004 to 2012 needs to be reprocessed to estimate the uncertainty for each ISAR SST_{skin} measurement. Second the AATSR $match - up$ processor needs to be updated to incorporate the QI method as well as to take advantage of the ISAR SST_{skin} record with uncertainties. With the updated AATSR $match - up$ processor the results of Wimmer *et al.* (2012)

can be revisited and the whole 2004 to 2012 $match-up$ archive reprocessed with the new methods.

Once the new $match-up$ processor is available the use of the ISAR record for validation is not limited to AATSR data and any satellite-derived SST data from the Bay of Biscay and the English Channel can be validated. The application to other SST products of the $match-up$ processor will also allow a more objective comparison of the SST product performances in which the derived uncertainty in of teh product is no longer dominated by the *in situ* measurement uncertainty or the $match-up$ quality (Corlett *et al.*, 2012).

Appendix A

Further three way uncertainty plots

This appendix shows the additional plots for three way uncertainty in chapter 5. The plots for the AATSR D3 data were shown in section 5.4.1, here are the plots for AATSR D2, N3 and N2 data.

First the plots for D2 data are shown, the histogram in figure A.1, the Q-Q plot for the unfiltered data in figure A.2 and the one standard deviation filtered data Q-Q plot in figure A.3. Next are the plots for the N3 data are shown again starting with the histogram in figure A.4, then the the Q-Q plot for the unfiltered data in figure A.5 and one standard deviation filtered data Q-Q plot in figure A.6. Finally the plots for N2 data are shown, the histogram in figure A.7, the Q-Q plot for the unfiltered data in figure A.8 and the one standard deviation filtered data Q-Q plot in figure A.9.

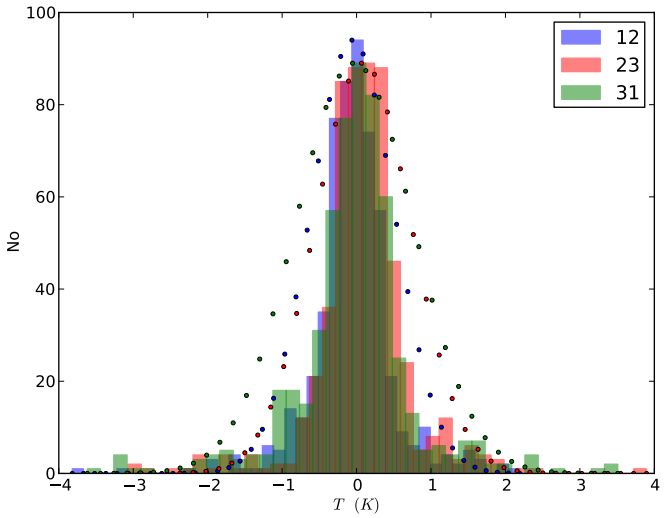


FIGURE A.1: Histogram of the data difference for AATSR D2 data. The coloured dots show a Gaussian distribution with the same mean and standard deviation as the data.

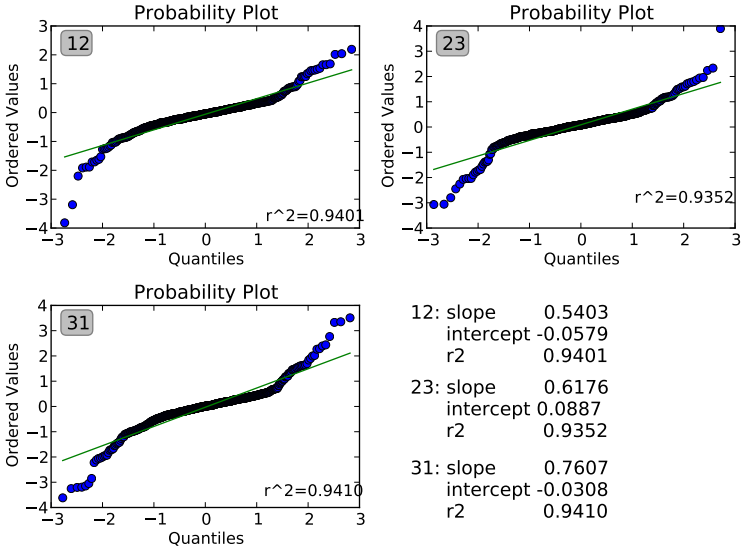


FIGURE A.2: QQ-plot of the data difference for AATSR D2 data.

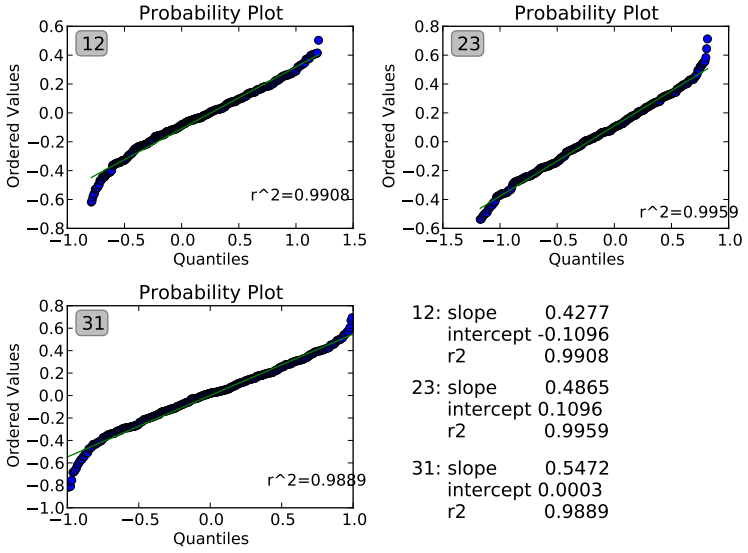


FIGURE A.3: QQ-plot of the data difference for AATSR D2 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot.

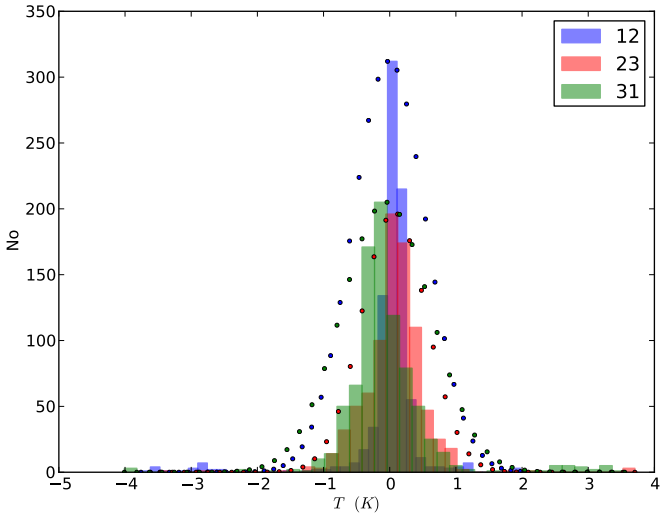


FIGURE A.4: Histogram of the data difference for AATSR N3 data. The coloured dots show a Gaussian distribution with the same mean and standard deviation as the data.

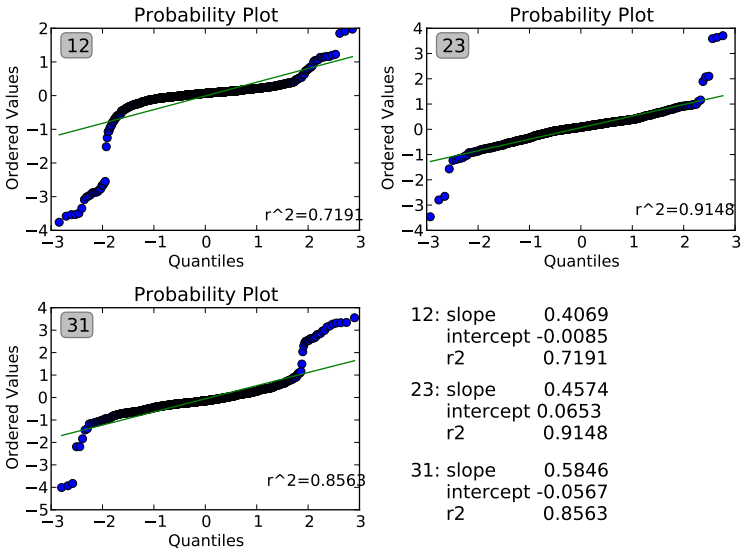


FIGURE A.5: QQ-plot of the data difference for AATSR N3 data.

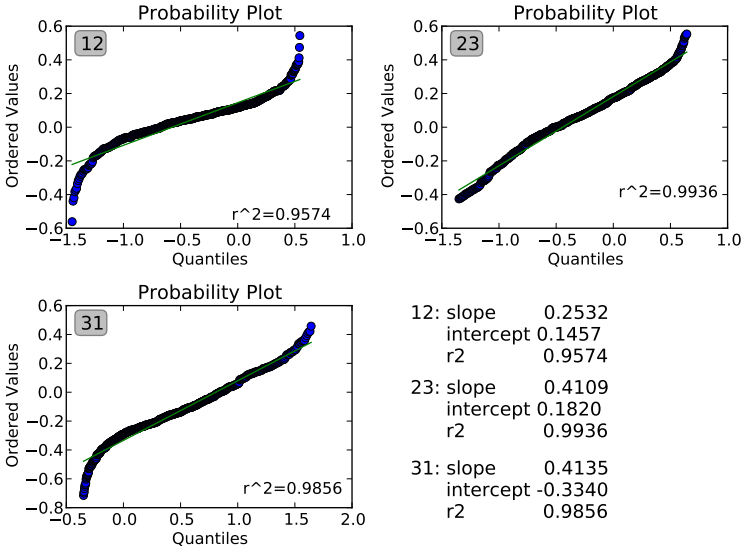


FIGURE A.6: QQ-plot of the data difference for AATSR N3 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot.

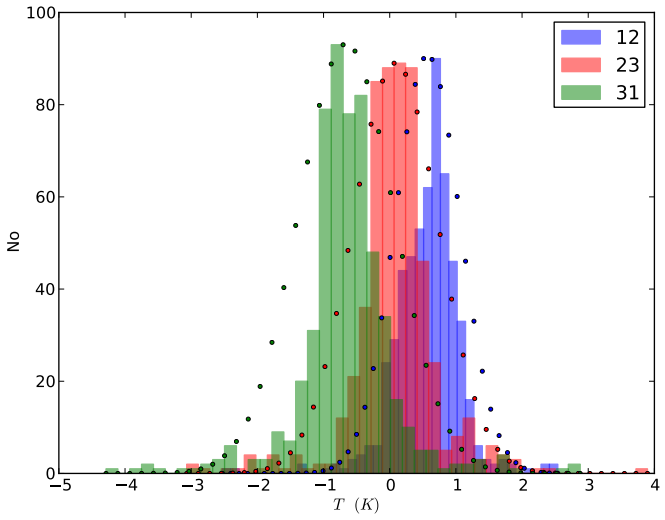


FIGURE A.7: Histogram of the data difference for AATSR N2 data. The coloured dots show a Gaussian distribution with the same mean and standard deviation as the data.

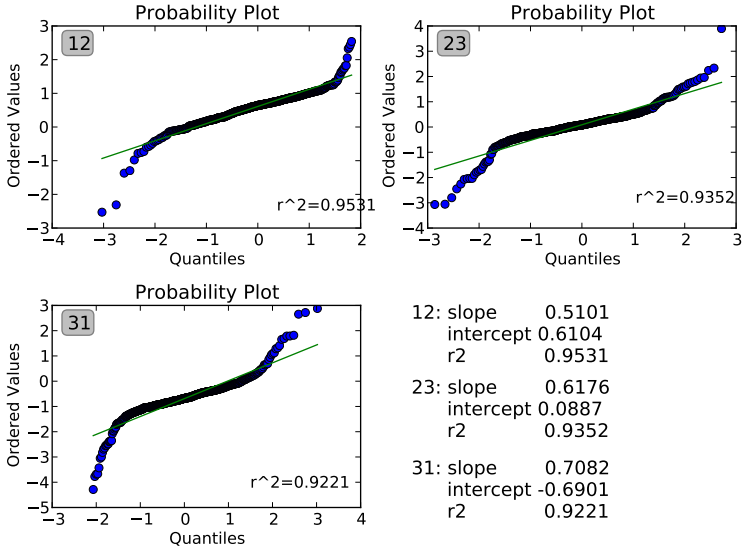


FIGURE A.8: QQ-plot of the data difference for AATSR N2 data.

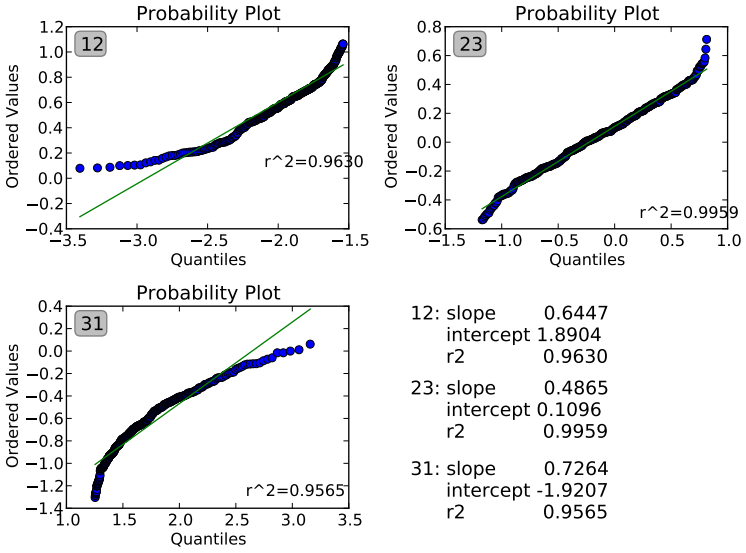


FIGURE A.9: QQ-plot of the data difference for AATSR N2 data. A one standard deviation filter has been applied to the data and a truncated Gaussian distribution was used for the plot.

Appendix B

Match-up QI results

In this appendix the additional results tables and plots for the QI can be found. Table 6.3 describes which of the data are presented in this appendix. To simplify the use of this appendix table 6.3 has been reproduced below.

Table B.1: Quality indicator results for 2004 to 2009 summary table. Table labels starting with B can be found in appendix B and table labels starting with a 6 can be found in section 6.6.

Data CV	Statistics	Table
0	Normal	B.2
	Robust	B.6
3	Normal	B.3
	Robust	B.7
4	Normal	B.4
	Robust	B.8
5	Normal	B.5
	Robust	6.4

First the statistics for the remaining AATSR CV values are listed in this appendix, followed by the $match - up$ plots as described in section 6.6 and grouped by CV value and statical method.

B.1 Statistics

TABLE B.2: Quality indicator results for 2004 to 2009 for AATSR data and normal statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	0.02	0.72	25642	486	273.92	298.39	0.02	359.98	0	0.32
q3	0.01	0.38	3824	254	273.92	298.39	0.17	359.80	0	0.32
q4	0.01	0.30	1633	156	273.92	298.39	0.23	359.85	0	0.28
q5	0.00	0.16	418	75	273.92	298.39	0.05	355.37	0	0.30
2 channel view										
q0	0.12	0.93	10298	230	273.92	297.56	0.15	359.98	0	0.32
q3	0.06	0.53	1225	106	273.92	297.56	0.63	358.90	0	0.32
q4	0.06	0.39	464	58	273.92	297.56	0.27	359.50	0	0.28
q5	0.04	0.22	106	29	273.92	297.56	0.55	292.35	0	0.30
3 channel view										
q0	-0.03	0.57	15349	256	277.11	298.39	0.02	359.90	0	0.32
q3	0.00	0.32	2601	148	277.11	298.39	0.17	359.80	0	0.32
q4	-0.01	0.25	1159	98	277.11	298.39	0.48	359.85	0	0.23
q5	-0.01	0.14	312	46	277.11	298.39	0.05	355.37	0	0.22
Nadir view										
q0	0.25	0.83	30064	486	273.92	298.39	0.02	359.98	0	0.32
q3	0.24	0.49	5602	254	273.92	298.39	0.02	359.80	0	0.32
q4	0.24	0.41	2335	156	273.92	298.39	0.18	359.85	0	0.31
q5	0.27	0.36	652	75	273.92	298.39	0.02	358.98	0	0.30
2 channel view										
q0	0.77	1.06	12528	230	273.92	297.56	0.07	359.98	0	0.32
q3	0.71	0.56	2022	106	273.92	297.56	0.18	358.90	0	0.32
q4	0.67	0.41	712	58	273.92	297.56	0.27	359.50	0	0.29
q5	0.67	0.33	205	29	273.92	297.56	0.40	358.98	0	0.21
3 channel view										
q0	-0.05	0.50	17520	256	277.11	298.39	0.02	359.98	0	0.32
q3	0.00	0.26	3578	148	277.11	298.39	0.02	359.80	0	0.32
q4	0.05	0.22	1607	98	277.11	298.39	0.18	359.85	0	0.31
q5	0.09	0.15	440	46	277.11	298.39	0.02	355.37	0	0.28

TABLE B.3: Quality indicator results for 2004 to 2009 for AATSR CV 3 data and normal statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	-1.34	1.22	260	52	273.92	298.39	4.12	359.12	0.00	0.31
q3	-1.40	0.70	8	4	273.92	298.39	16.40	251.47	0.00	0.24
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
2 channel view										
q0	-1.34	1.22	260	52	273.92	297.56	4.12	359.12	0.00	0.31
q3	-1.40	0.70	8	4	273.92	297.56	16.40	251.47	0.00	0.24
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
3 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
Nadir view										
q0	0.45	1.12	259	52	273.92	298.39	4.12	359.12	0.00	0.31
q3	0.51	0.39	31	4	273.92	298.39	9.05	327.72	0.00	0.23
q4	0.57	0.31	3	1	273.92	298.39	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0.0	0.0	0.0	0.0	0.0	0.0
2 channel view										
q0	0.45	1.12	259	52	273.92	297.56	4.12	359.12	0.00	0.31
q3	0.51	0.39	31	4	273.92	297.56	9.05	327.72	0.00	0.23
q4	0.57	0.31	3	1	273.92	297.56	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0.0	0.0	0.0	0.0	0.0	0.0
3 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0

TABLE B.4: Quality indicator results for 2004 to 2009 for AATSR CV 4 data and normal statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	-0.77	0.74	726	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	-0.20	1.22	58	26	273.92	298.39	8.28	351.62	0.00	0.23
q4	-0.65	0.24	11	6	273.92	298.39	5.10	295.12	0.00	0.02
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
2 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
3 channel view										
q0	-0.77	0.74	726	123	277.11	298.39	0.10	359.62	0.00	0.31
q3	-0.20	1.22	58	26	277.11	298.39	8.28	351.62	0.00	0.23
q4	-0.65	0.24	11	6	277.11	298.39	5.10	295.12	0.00	0.02
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
Nadir view										
q0	-0.13	0.68	723	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	0.46	1.35	95	26	273.92	298.39	0.10	342.50	0.00	0.28
q4	-0.11	0.25	29	6	273.92	298.39	1.43	327.07	0.00	0.23
q5	-0.21	0.19	6	1	273.92	298.39	0.0	0.0	0.0	0.0
2 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
3 channel view										
q0	-0.13	0.68	723	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	0.46	1.35	95	26	273.92	298.39	0.10	342.50	0.00	0.28
q4	-0.11	0.25	29	6	273.92	298.39	1.43	327.07	0.00	0.23
q5	-0.21	0.19	6	1	273.92	298.39	0.0	0.0	0.0	0.0

TABLE B.5: Quality indicator results for 2004 to 2009 for AATSR CV 5 data and normal statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	0.04	0.66	24022	442	273.92	298.39	0.0167	360.0	0	0.32
q3	0.02	0.36	3728	252	273.92	298.39	0.1667	359.8	0	0.32
q4	0.01	0.29	1614	156	273.92	298.39	0.23	359.85	0	0.28
q5	0.00	0.16	418	75	273.92	298.39	0.05	355.37	0	0.30
2 channel view										
q0	0.11	0.87	9508	208	273.92	297.56	0.15	359.98	0	0.32
q3	0.05	0.50	1192	104	273.92	297.56	0.63	358.90	0	0.32
q4	0.05	0.39	460	58	273.92	297.56	0.27	359.50	0	0.28
q5	0.04	0.22	106	29	273.92	297.56	0.55	292.35	0	0.30
3 channel view										
q0	0.01	0.53	14510	234	277.11	298.39	0.02	359.90	0	0.32
q3	0.01	0.30	2534	148	277.11	298.39	0.17	359.80	0	0.32
q4	-0.01	0.24	1145	98	277.11	298.39	0.48	359.85	0	0.23
q5	-0.01	0.14	312	46	277.11	298.39	0.05	355.37	0	0.22
Nadir view										
q0	0.28	0.75	24045	442	273.92	298.39	0.02	359.98	0	0.32
q3	0.24	0.47	4870	252	273.92	298.39	0.13	359.80	0	0.32
q4	0.24	0.40	2149	156	273.92	298.39	0.18	359.85	0	0.31
q5	0.27	0.35	628	75	273.92	298.39	0.02	358.98	0	0.30
2 channel view										
q0	0.81	0.89	9530	208	273.92	297.56	0.15	359.98	0	0.32
q3	0.70	0.50	1654	104	273.92	297.56	0.18	358.90	0	0.32
q4	0.67	0.42	641	58	273.92	297.56	0.27	359.50	0	0.29
q5	0.67	0.33	194	29	273.92	297.56	0.55	358.98	0	0.21
3 channel view										
q0	-0.02	0.46	14499	234	277.11	298.39	0.02	359.90	0	0.32
q3	0.01	0.25	3206	148	277.11	298.39	0.13	359.80	0	0.32
q4	0.06	0.22	1496	98	277.11	298.39	0.18	359.85	0	0.31
q5	0.10	0.14	426	46	277.11	298.39	0.02	355.37	0	0.28

TABLE B.6: Quality indicator results for 2004 to 2009 for AATSR all data and robust statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	0.02	0.47	26359	486	273.92	298.39	0.02	359.98	0	0.32
q3	0.02	0.28	3901	254	273.92	298.39	0.17	359.80	0	0.32
q4	0.00	0.21	1650	156	273.92	298.39	0.23	359.85	0	0.28
q5	0.00	0.15	425	75	273.92	298.39	0.05	355.37	0	0.30
2 channel view										
q0	0.10	0.59	10619	230	273.92	297.56	0.02	359.98	0	0.32
q3	0.06	0.37	1267	106	273.92	297.56	0.63	358.90	0	0.32
q4	0.04	0.31	473	58	273.92	297.56	0.27	359.50	0	0.28
q5	0.03	0.25	110	29	273.92	297.56	0.55	317.03	0	0.30
3 channel view										
q0	-0.02	0.40	15740	256	277.11	298.39	0.02	359.90	0	0.32
q3	0.00	0.24	2634	148	277.11	298.39	0.17	359.80	0	0.32
q4	-0.02	0.18	1177	98	277.11	298.39	0.23	359.85	0	0.24
q5	-0.01	0.13	315	46	277.11	298.39	0.05	355.37	0	0.22
Nadir view										
q0	0.24	0.55	30921	486	273.92	298.39	0.02	359.98	0	0.32
q3	0.23	0.34	5708	254	273.92	298.39	0.02	359.80	0	0.32
q4	0.22	0.30	2355	156	273.92	298.39	0.18	359.85	0	0.31
q5	0.24	0.30	654	75	273.92	298.39	0.02	358.98	0	0.30
2 channel view										
q0	0.71	1.22	12934	230	273.92	297.56	0.02	359.98	0	0.32
q3	0.69	1.10	2080	106	273.92	297.56	0.18	358.90	0	0.32
q4	0.68	1.08	723	58	273.92	297.56	0.27	359.50	0	0.29
q5	0.69	1.04	207	29	273.92	297.56	0.40	358.98	0	0.21
3 channel view										
q0	-0.04	0.31	17987	256	277.11	298.39	0.02	359.98	0	0.32
q3	0.02	0.21	3628	148	277.11	298.39	0.02	359.80	0	0.32
q4	0.06	0.19	1632	98	277.11	298.39	0.18	359.85	0	0.31
q5	0.10	0.19	447	46	277.11	298.39	0.02	355.37	0	0.30

TABLE B.7: Quality indicator results for 2004 to 2009 for AATSR CV 3 data and robust statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	-1.29	1.96	267	52	273.92	298.39	4.12	359.12	0.00	0.31
q3	-1.29	1.99	8	4	273.92	298.39	16.40	251.47	0.00	0.24
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
2 channel view										
q0	-1.29	1.96	267	52	273.92	298.39	4.12	359.12	0.00	0.31
q3	-1.29	1.99	8	4	273.92	298.39	16.40	251.47	0.00	0.24
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
3 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
Nadir view										
q0	0.50	1.19	267	52	273.92	298.39	4.12	359.12	0.00	0.31
q3	0.50	0.83	32	4	273.92	298.39	9.05	327.72	0.00	0.24
q4	0.0	1.07	3	1	273.92	298.39	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	273.92	298.39	0.0	0.0	0.0	0.0
2 channel view										
q0	0.50	1.19	267	52	273.92	298.39	4.12	359.12	0.00	0.31
q3	0.50	0.83	32	4	273.92	298.39	9.05	327.72	0.00	0.24
q4	0.0	1.07	3	1	273.92	298.39	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	273.92	298.39	0.0	0.0	0.0	0.0
3 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0

TABLE B.8: Quality indicator results for 2004 to 2009 for AATSR CV 4 data and robust statistics.

Quality	Mean	Standard deviation	No. of Matches	No. of Overpasses	Min temp, K	Max temp, K	Min time diff, min	Max time diff, min	Min distance offset, deg	Max distance offset, deg
Dual view										
q0	-0.76	1.25	752	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	-0.50	0.98	59	26	273.92	298.39	8.28	351.62	0.00	0.23
q4	0.0	0.83	11	6	273.92	298.39	5.10	295.12	0.00	0.02
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
2 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
3 channel view										
q0	-0.76	1.25	752	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	-0.50	0.98	59	26	273.92	298.39	8.28	351.62	0.00	0.23
q4	0.0	0.83	11	6	273.92	298.39	5.10	295.12	0.00	0.02
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.
Nadir view										
q0	-0.11	0.52	752	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	0.09	0.42	95	26	273.92	298.39	0.10	342.50	0.00	0.28
q4	-0.11	0.31	29	6	273.92	298.39	1.43	327.07	0.00	0.23
q5	-0.21	0.28	6	1	273.92	298.39	0.0	0.0	0.0	0.0
2 channel view										
q0	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q3	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q4	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
q5	0.00	0.00	0	0	0	0	0.0	0.0	0.0	0.0
3 channel view										
q0	-0.11	0.52	752	123	273.92	298.39	0.10	359.62	0.00	0.31
q3	0.09	0.42	95	26	273.92	298.39	0.10	342.50	0.00	0.28
q4	-0.11	0.31	29	6	273.92	298.39	1.43	327.07	0.00	0.23
q5	-0.21	0.28	6	1	273.92	298.39	0.0	0.0	0.0	0.0

B.2 Plots

B.2.1 AATSR N1 data with normal statistics

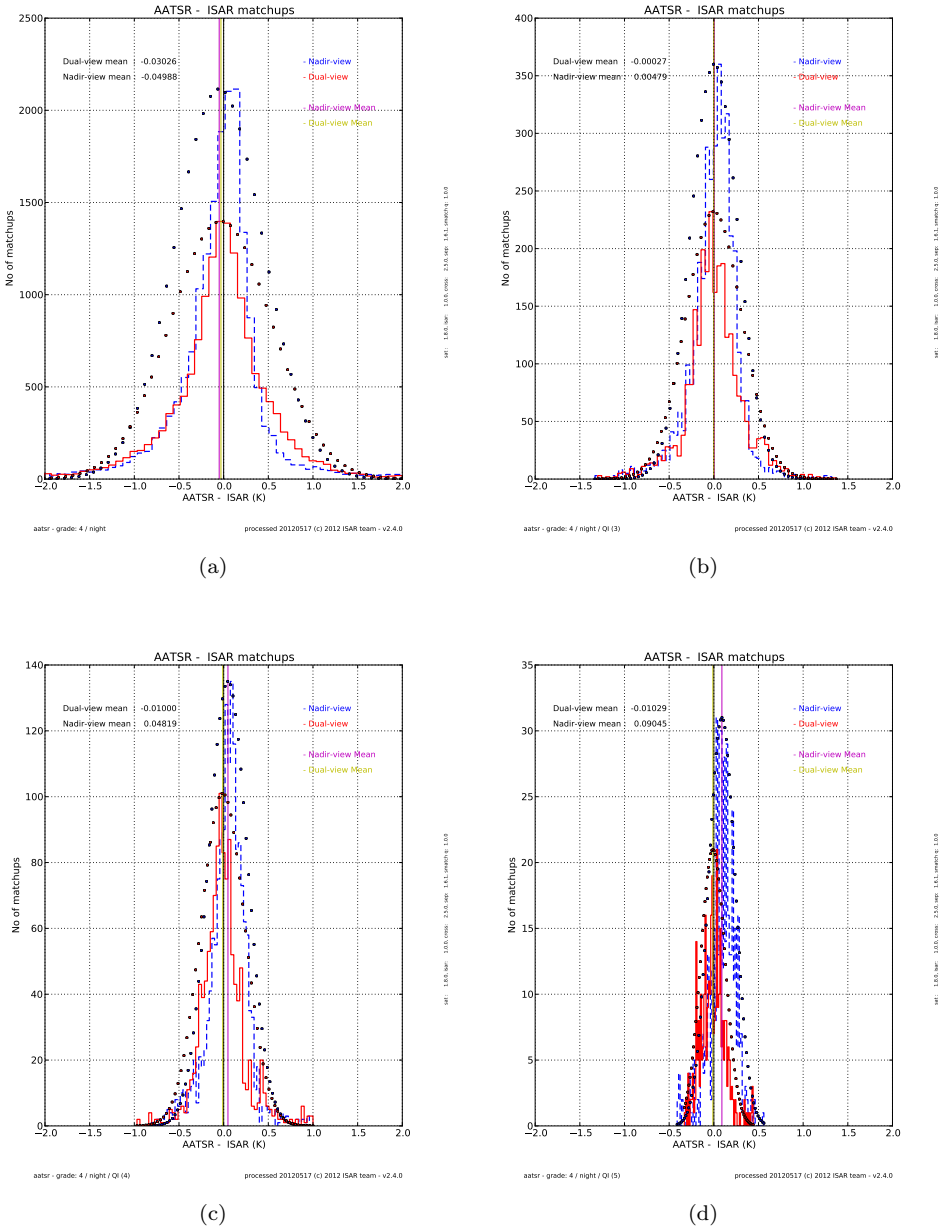


FIGURE B.1: QI histograms for N1 and normal statistics for night time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

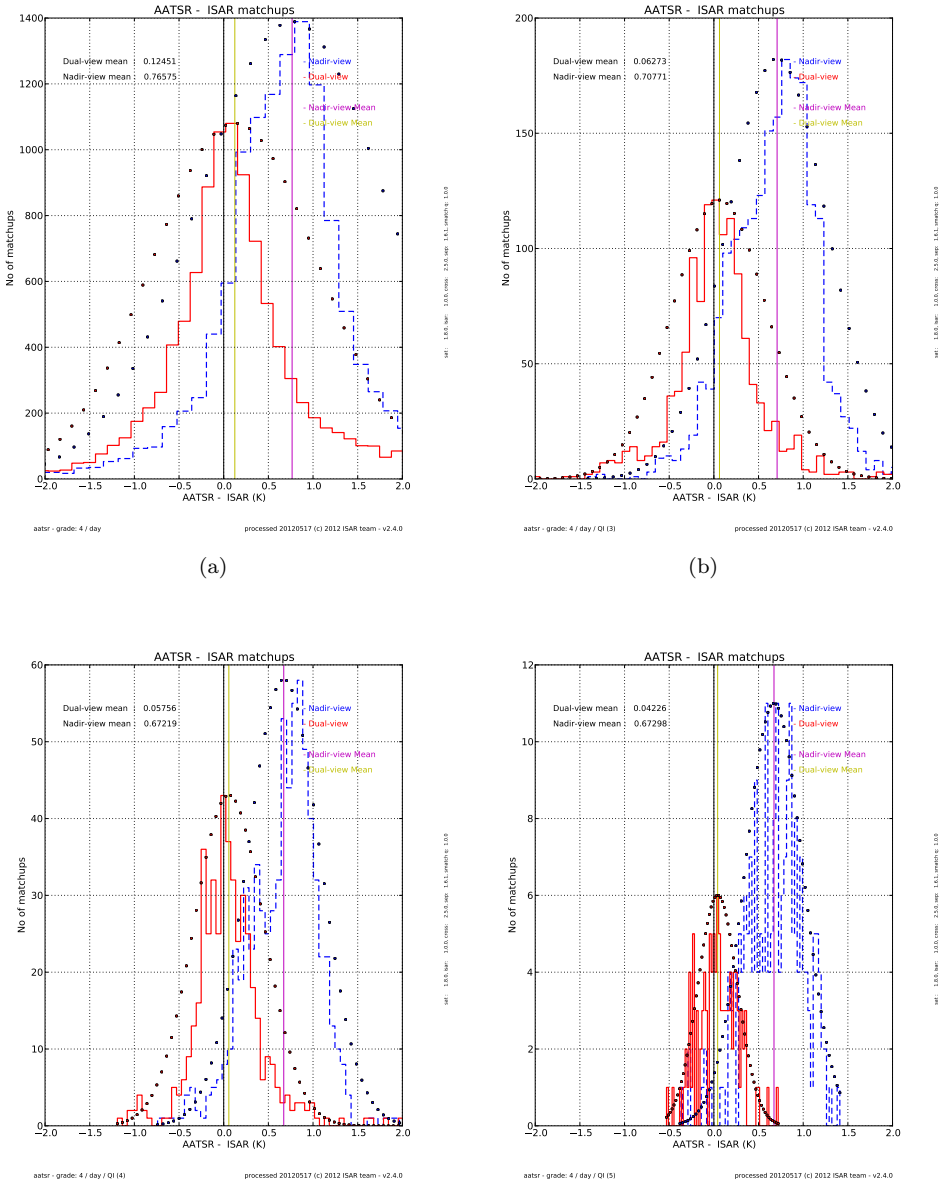


FIGURE B.2: QI histograms for and normal statistics for day time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

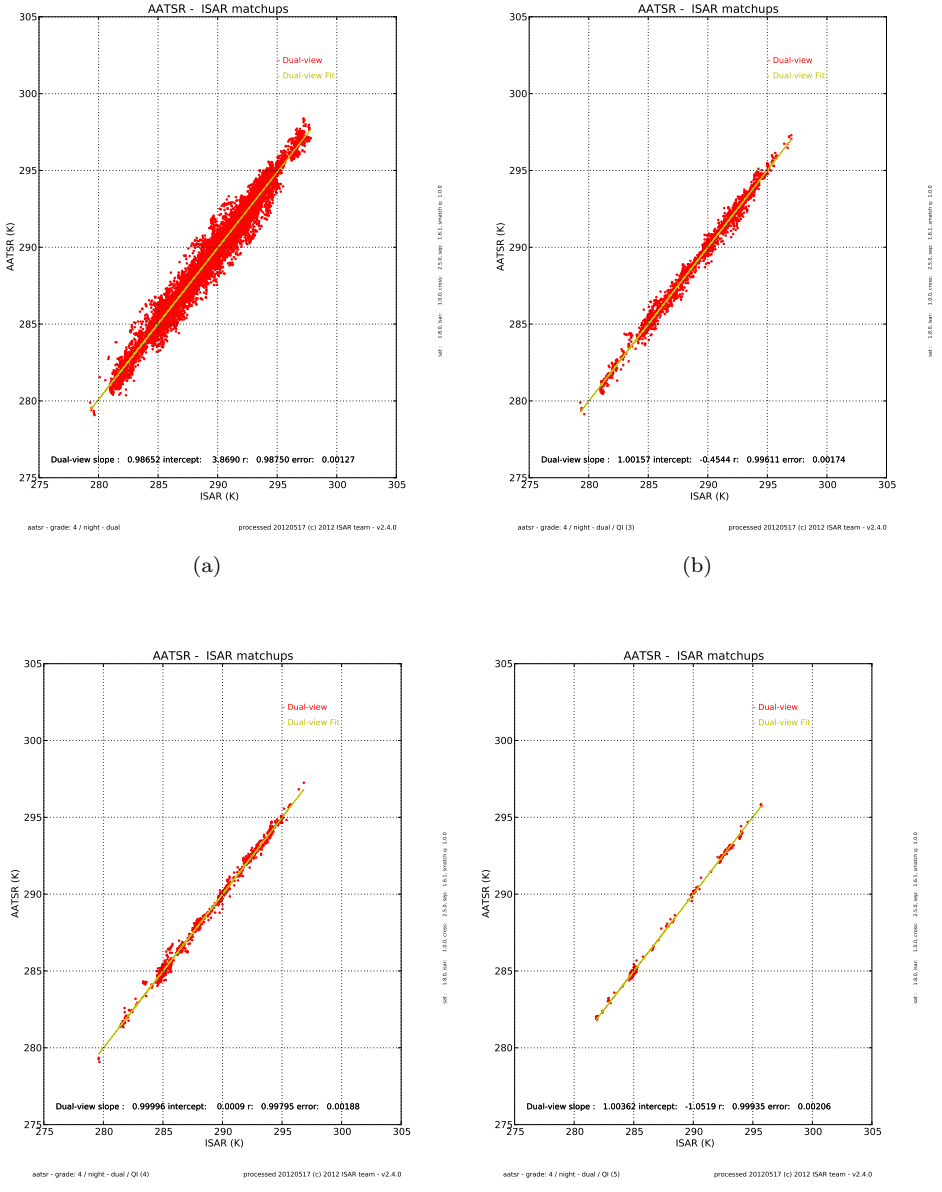


FIGURE B.3: QI scatter plot for N1 and normal statistics for dual view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

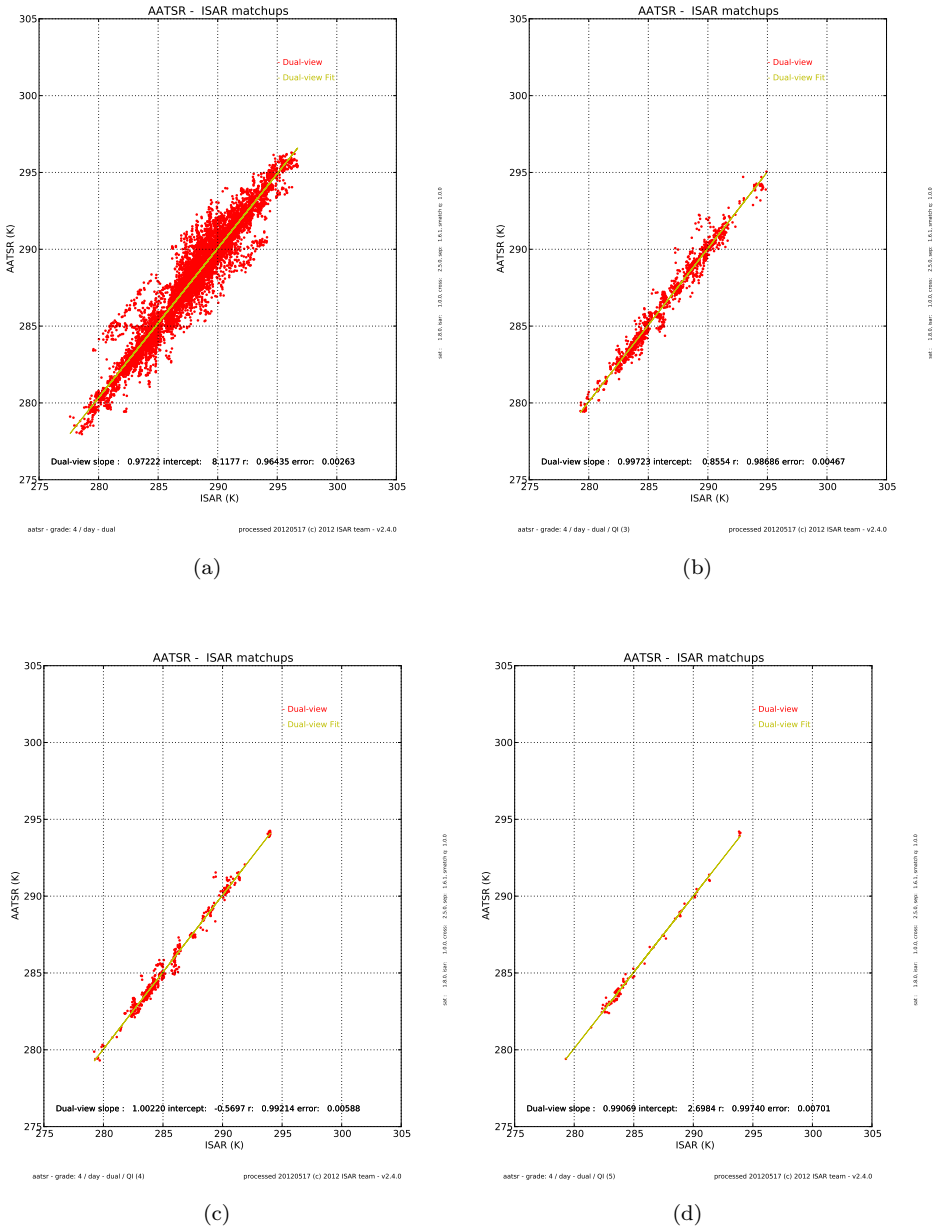


FIGURE B.4: QI scatter plot for N1 and normal statistics for dual view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

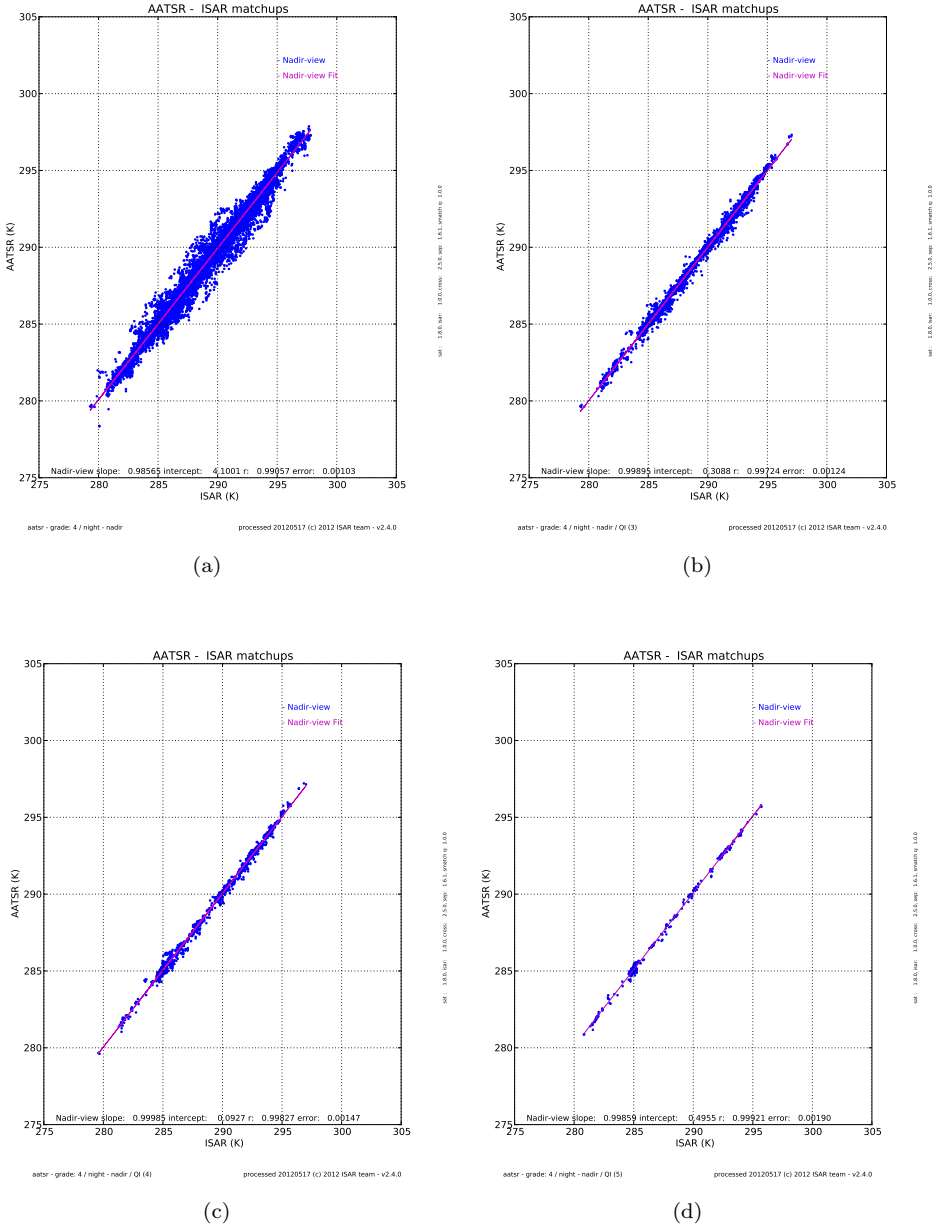


FIGURE B.5: QI scatter plot for N1 and normal statistics for nadir view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

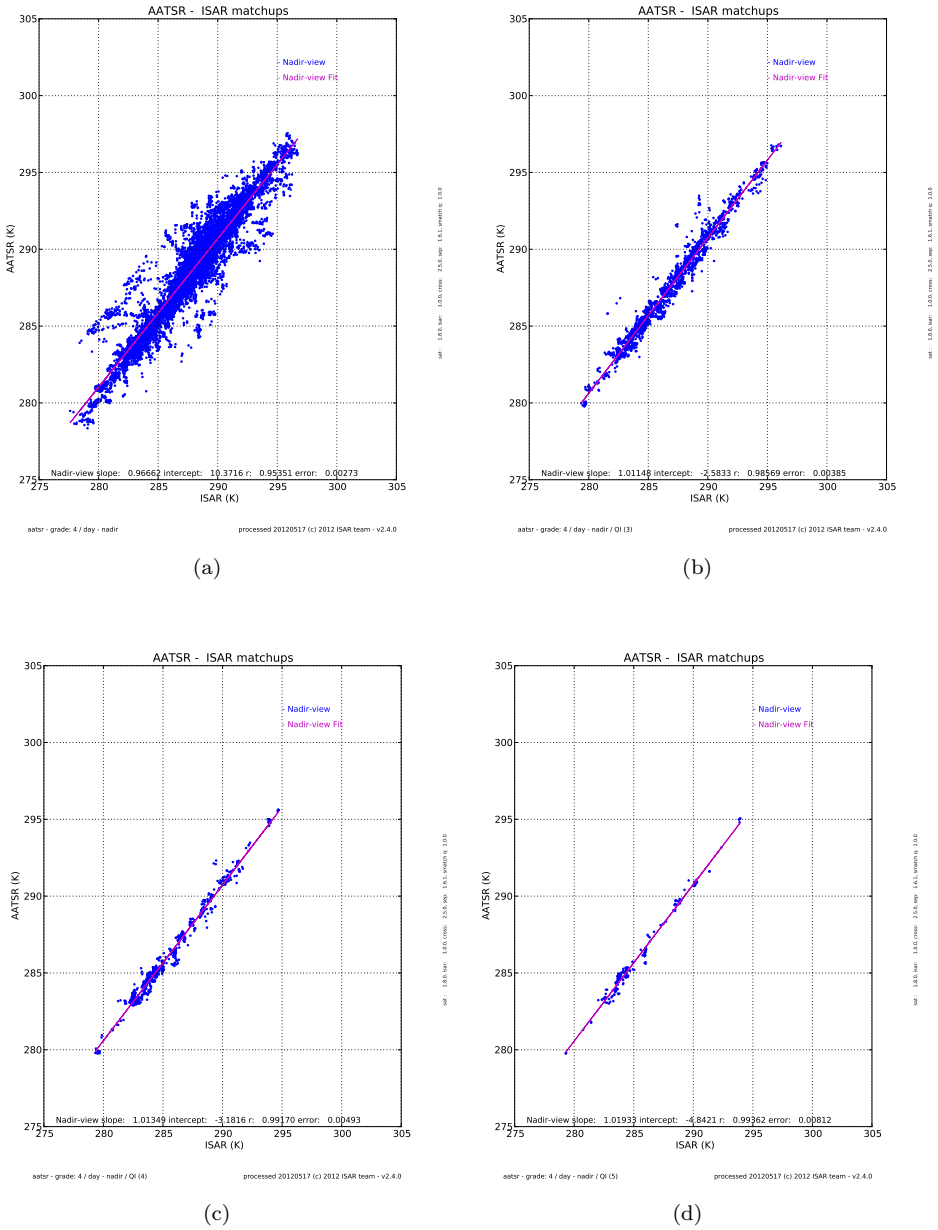


FIGURE B.6: QI scatter plot for N1 and normal statistics for nadir view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

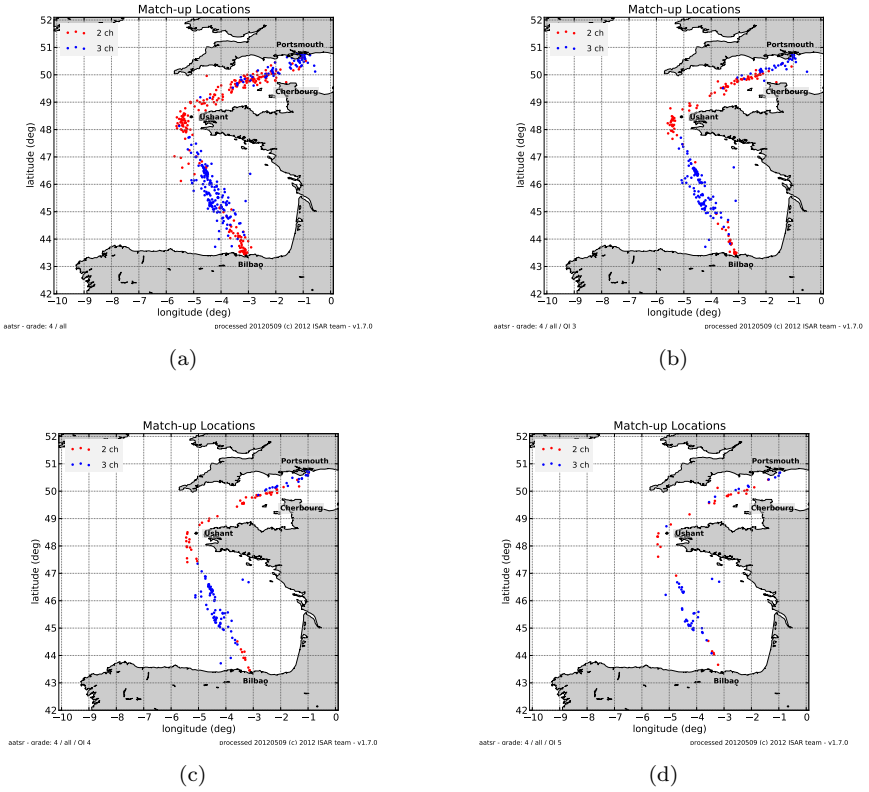
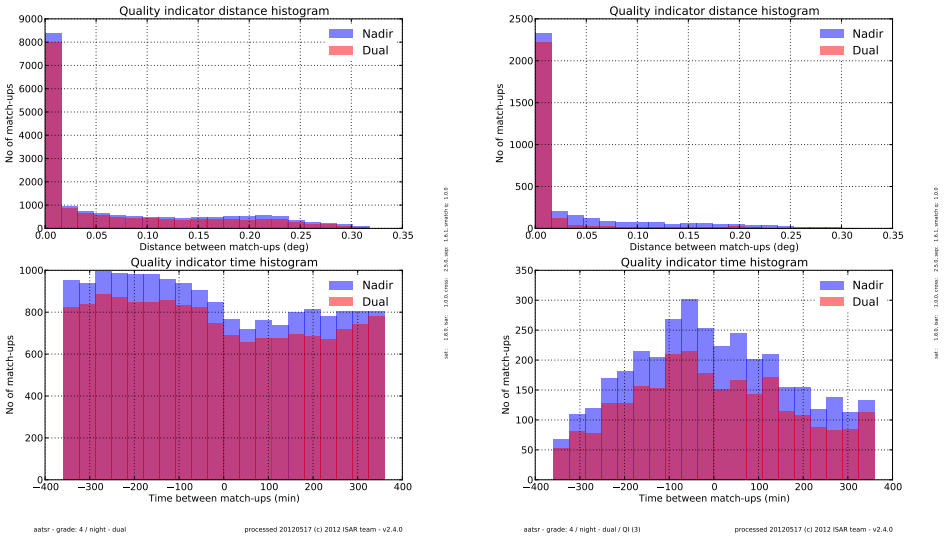
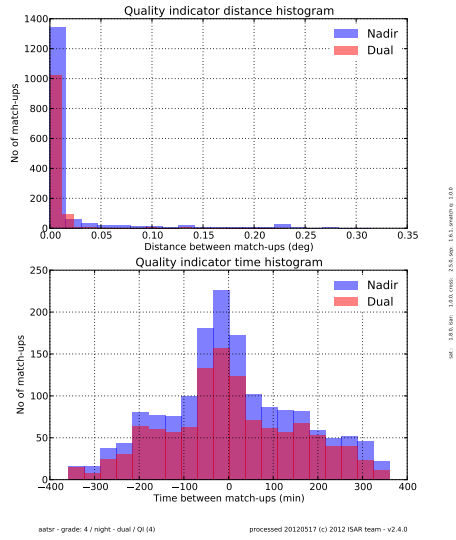


FIGURE B.7: QI location plot for N1 and normal statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; (c) QI 4 location plot; and, (d) QI 5 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue

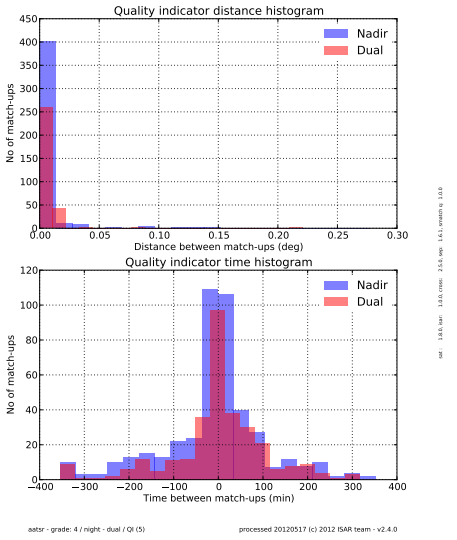


(a)

(b)

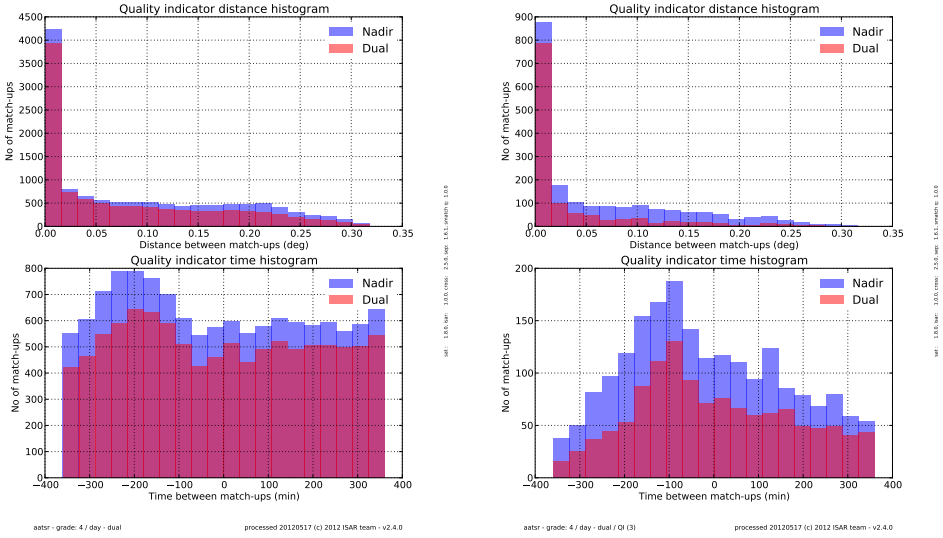


(c)



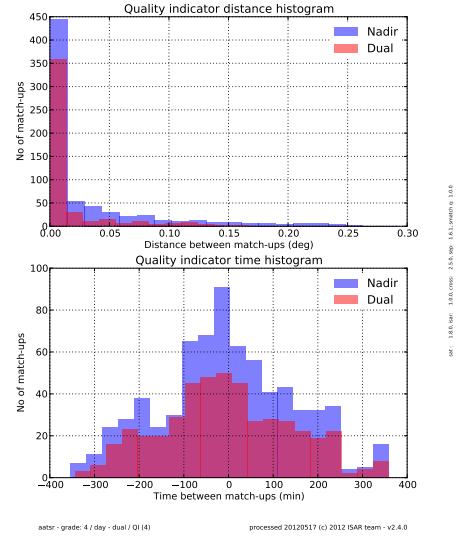
(d)

FIGURE B.8: QI time and distance plot for N1 and normal statistics for for night time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue

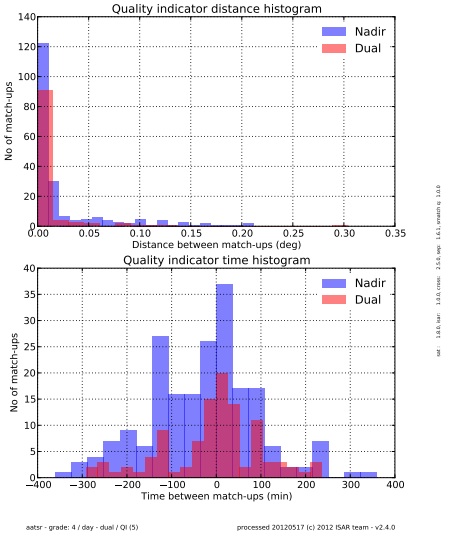


(a)

(b)



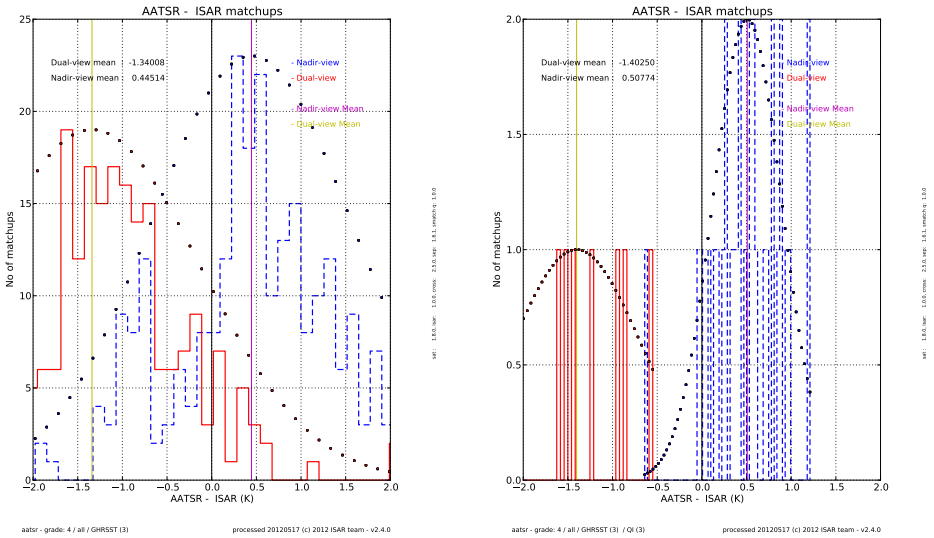
(c)



(d)

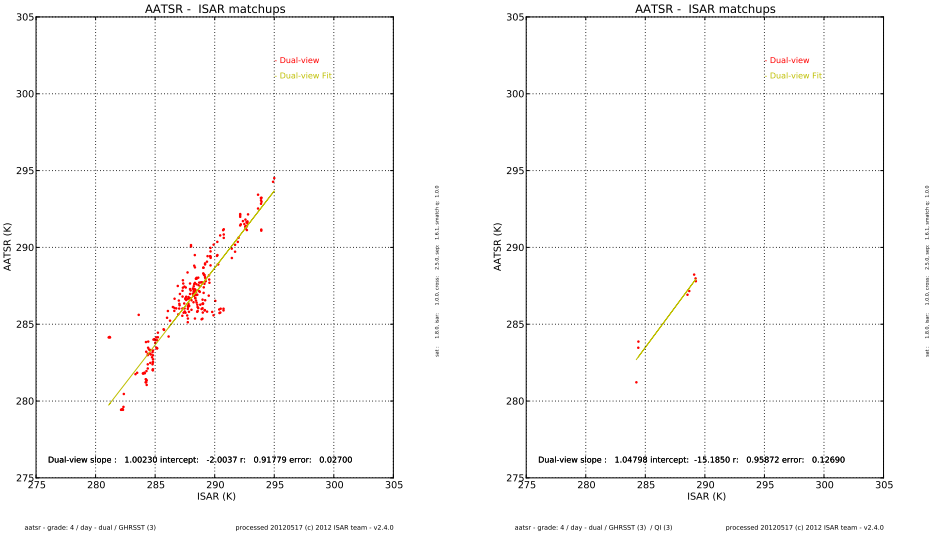
FIGURE B.9: QI time and distance plot for N1 and normal statistics for for day time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue

B.2.2 AATSR CV 3 data with normal statistics



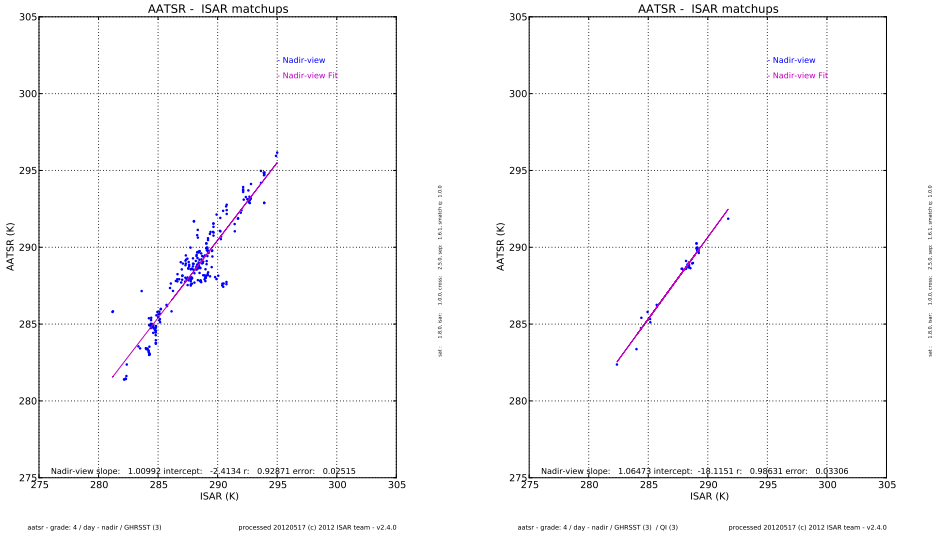
(a) (b)

FIGURE B.10: QI histograms for and normal statistics for day time data: (a) QI 0 histogram; (a) QI 3 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like. No histograms for QI 4 and 5 were produced as there was no data for QI 4 and 5.

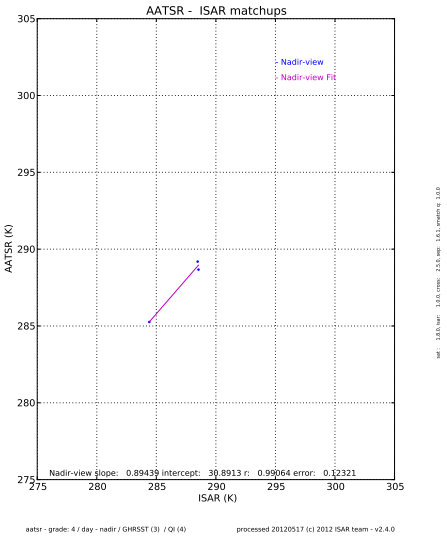


(a) (b)

FIGURE B.11: QI scatter plot for ghrsst3 and normal statistics for dual view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; No QI 4 and 5 data is available.

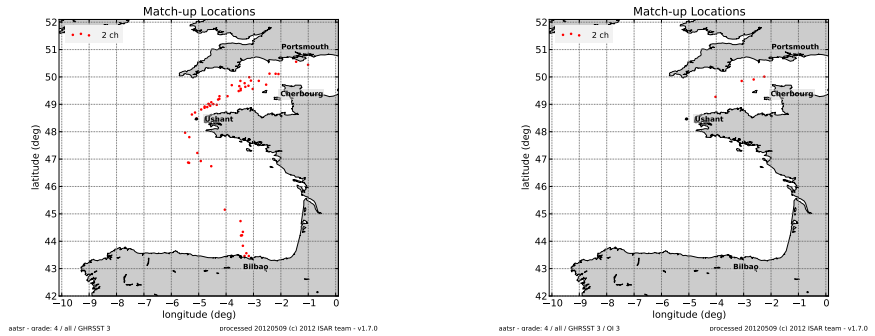


(a) (b)



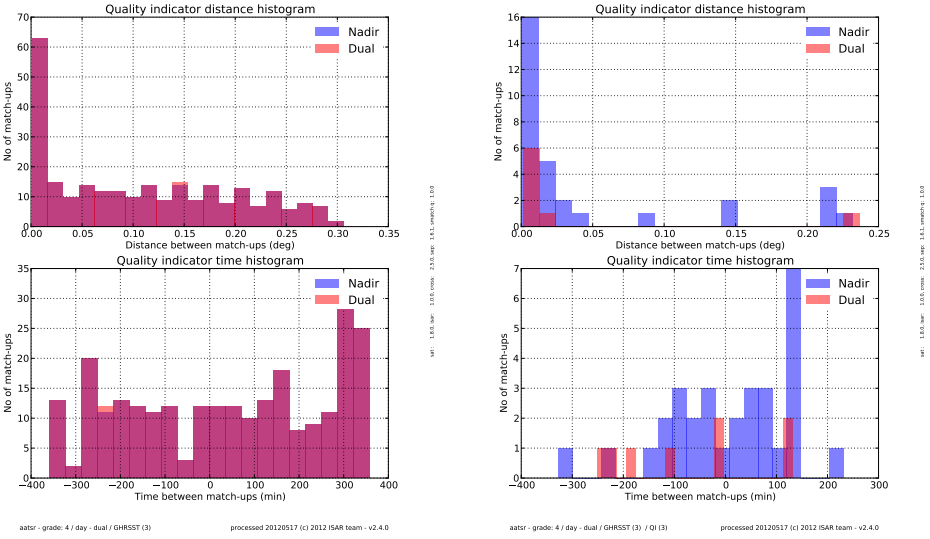
(c)

FIGURE B.12: QI scatter plot for ghrsst3 and normal statistics for nadir view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, No QI 5 data were produced.



(a) (b)

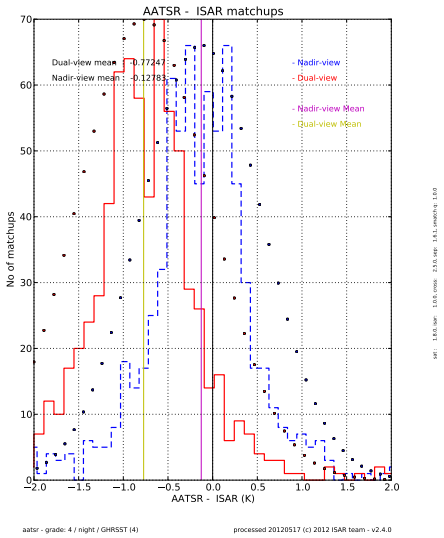
FIGURE B.13: QI location plot for ghrsst3 and normal statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue. No QI 4 or 5 data were produced.



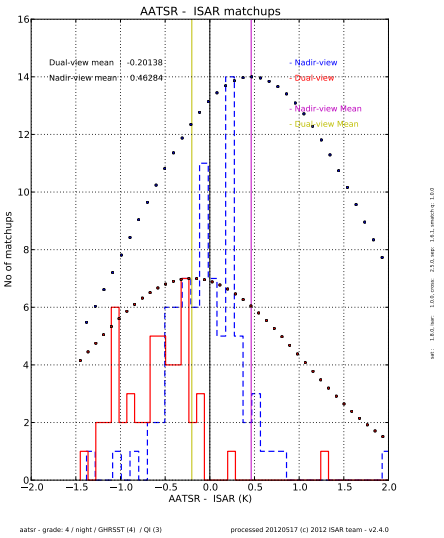
(a) (b)

FIGURE B.14: QI time and distance plot for CV 3 and normal statistics for for day time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue. No QI 4 and 5 data available.

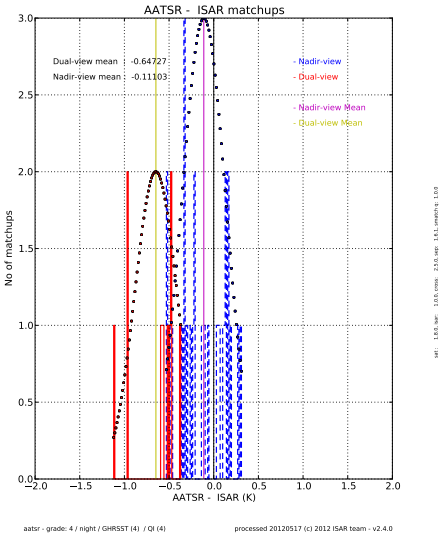
B.2.3 AATSR CV 4 data with normal statistics



(a)



(b)



(c)

FIGURE B.15: QI histograms for CV 4 and normal statistics for night time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; The red histogram shows dual view $match-up$ data, the dashed blue histogram shows nadir view only $match-up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like. No QI 5 data was produced

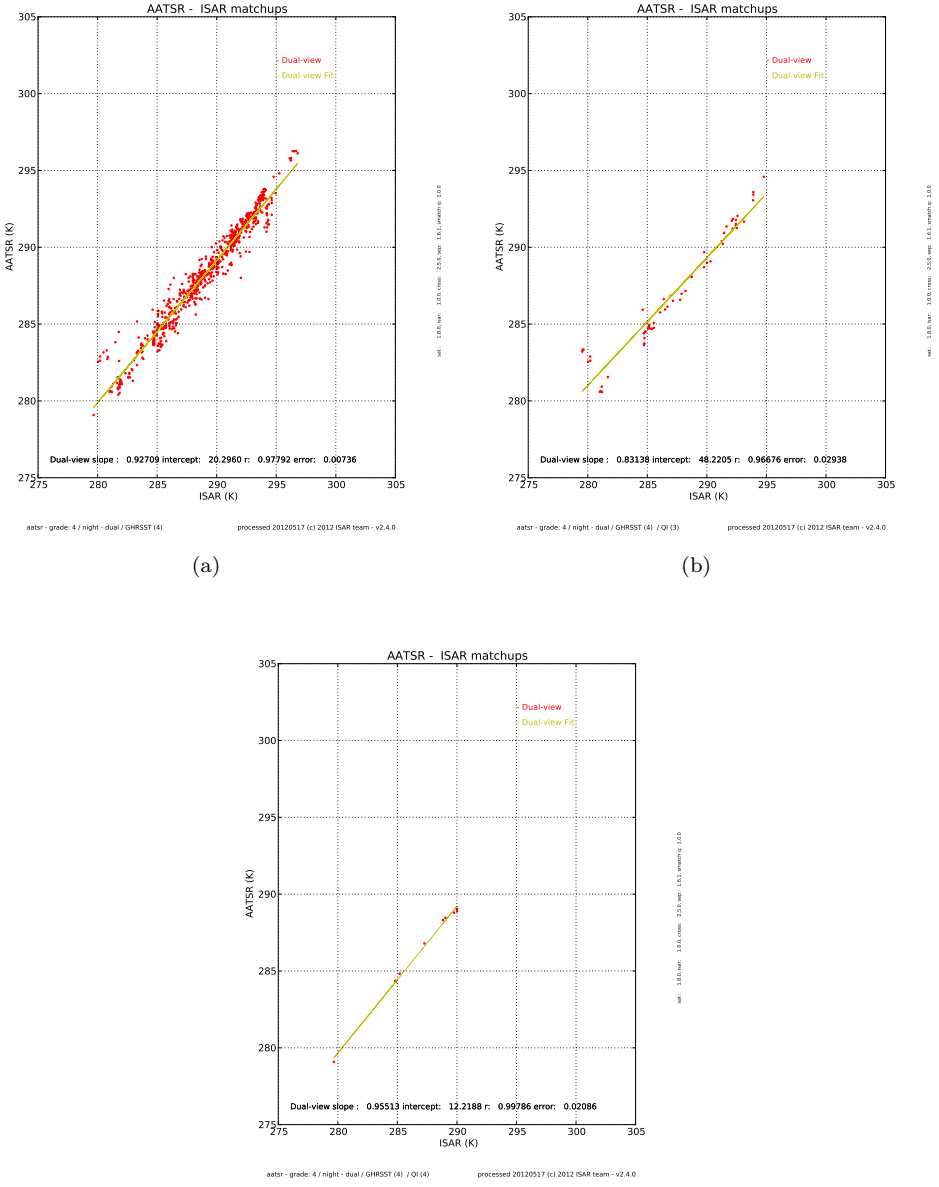
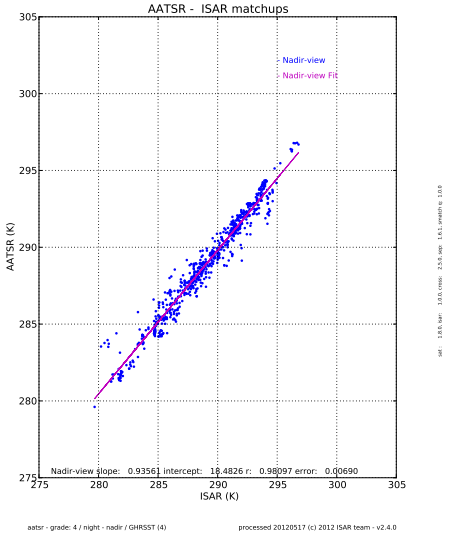
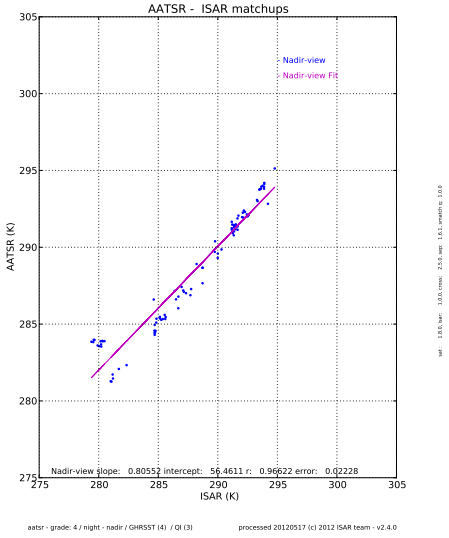


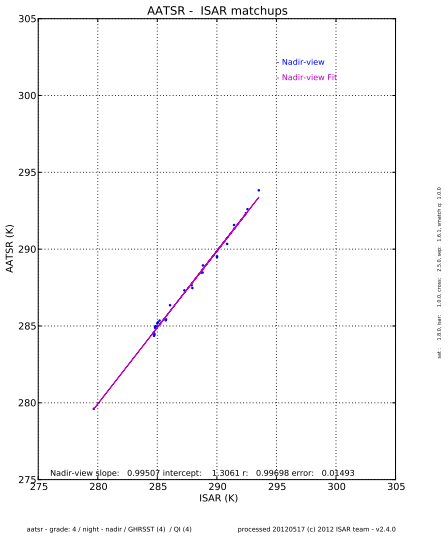
FIGURE B.16: QI scatter plot for CV 4 and normal statistics for dual view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot;No QI 5 data were produced.



(a)

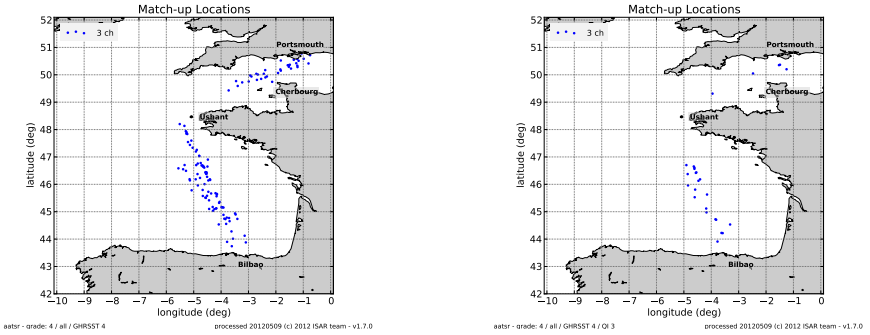


(b)



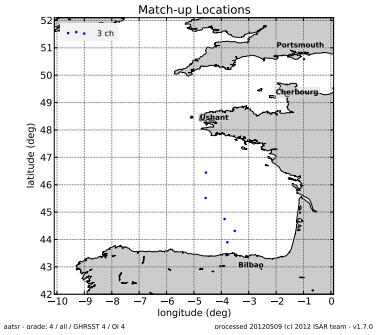
(c)

FIGURE B.17: QI scatter plot for CV 4 and normal statistics for nadir view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot;No QI 5 data were produced



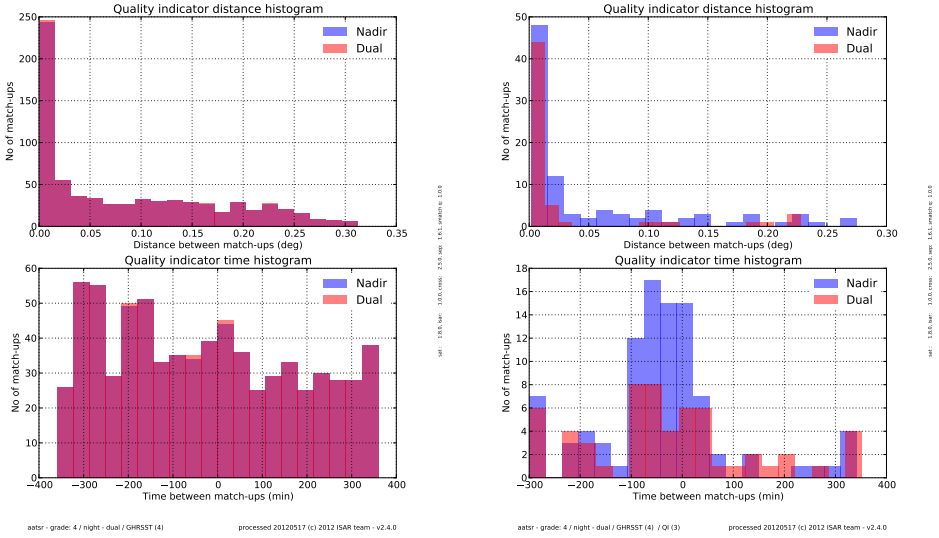
(a)

(b)

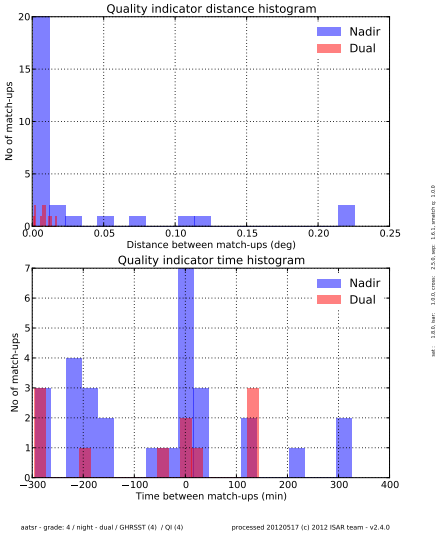


(c)

FIGURE B.18: QI location plot for CV 4 and normal statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; (c) QI 4 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue. No QI 5 data were produced.



(a) (b)



(c)

FIGURE B.19: QI time and distance plot for CV 4 and normal statistics for for night time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue. No QI 5 data were produced.

B.2.4 AATSR CV 5 data with normal statistics

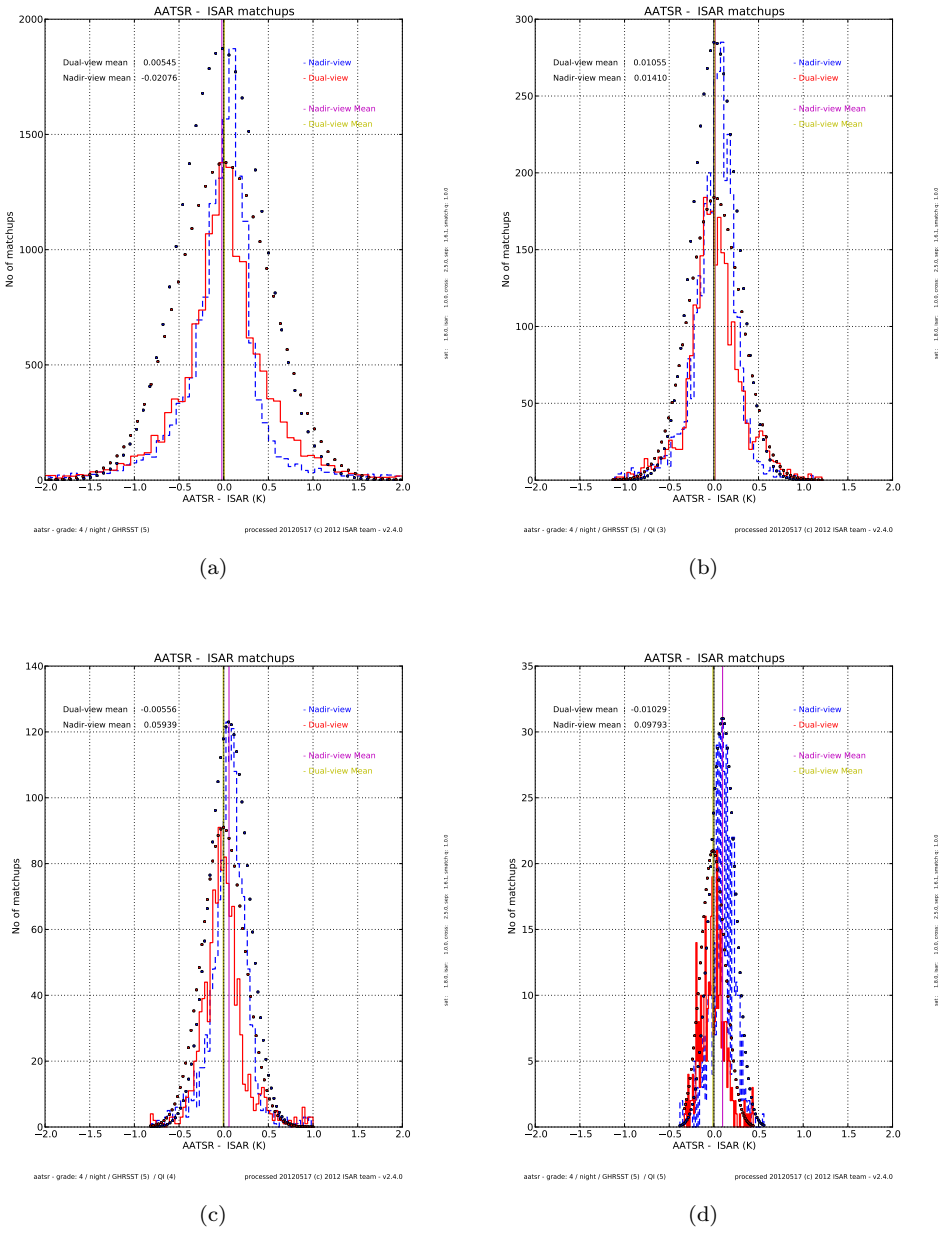


FIGURE B.20: QI histograms for CV 5 and normal statistics for night time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view $match-up$ data, the dashed blue histogram shows nadir view only $match-up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

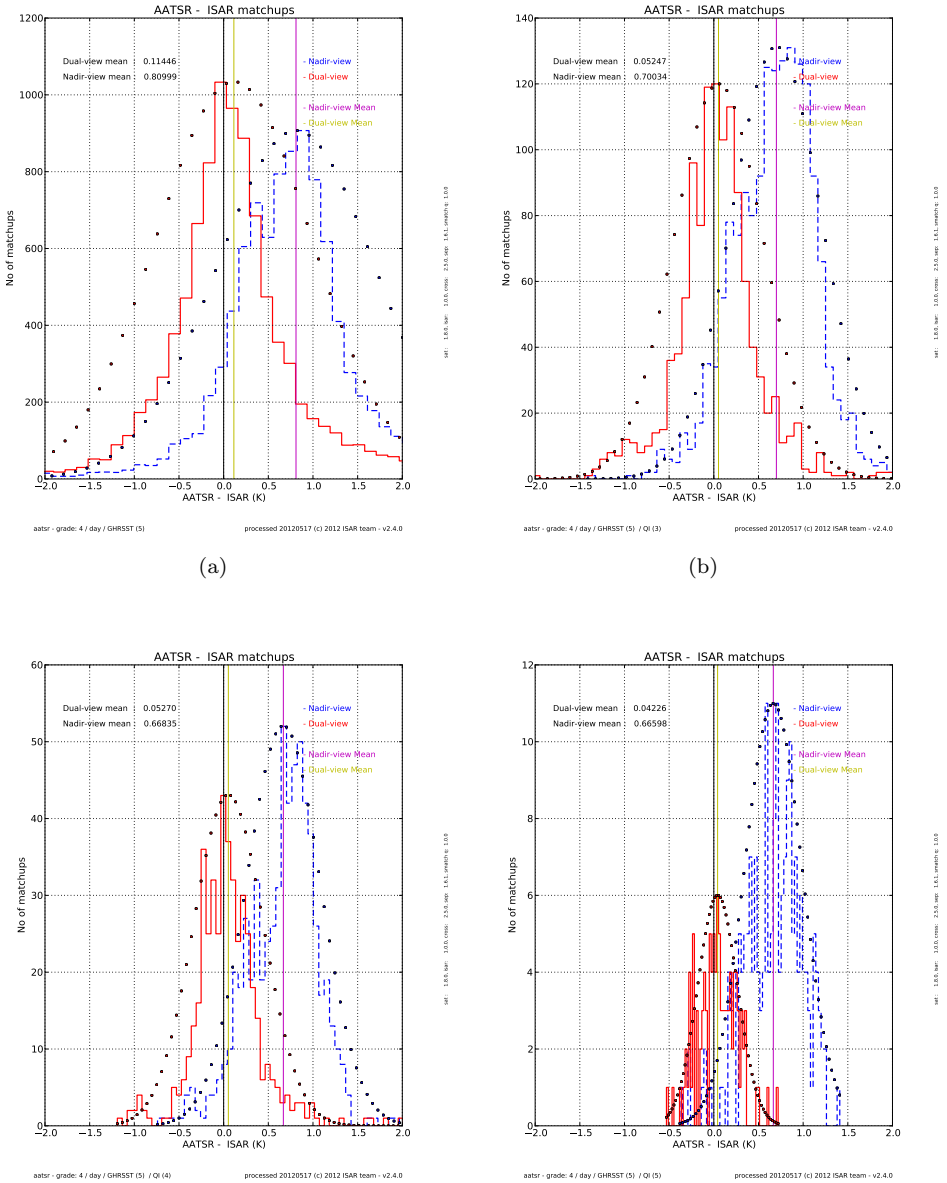


FIGURE B.21: QI histograms for CV 5 and normal statistics for day time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

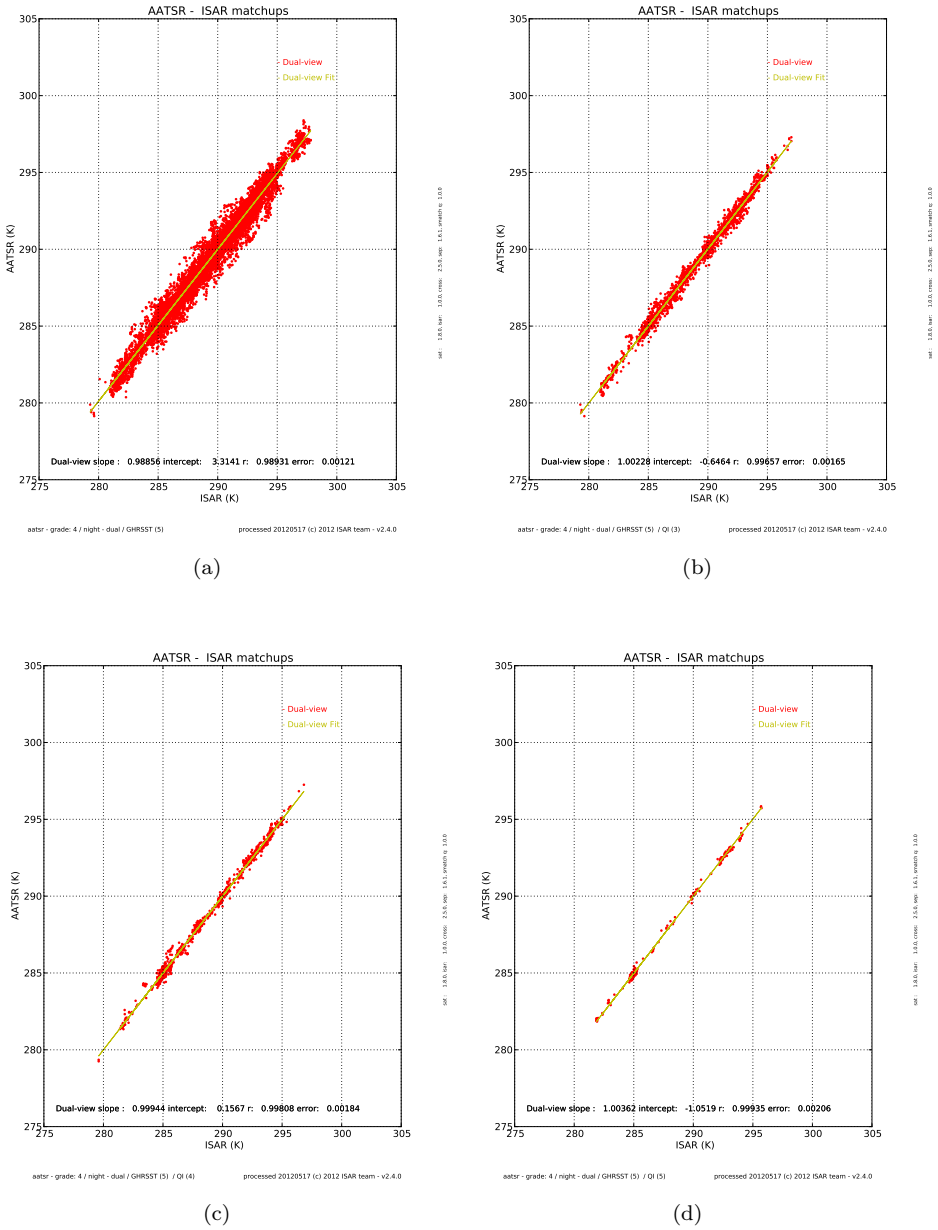


FIGURE B.22: QI scatter plot for CV 5 and normal statistics for dual view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, **??** QI 5 scatter plot;

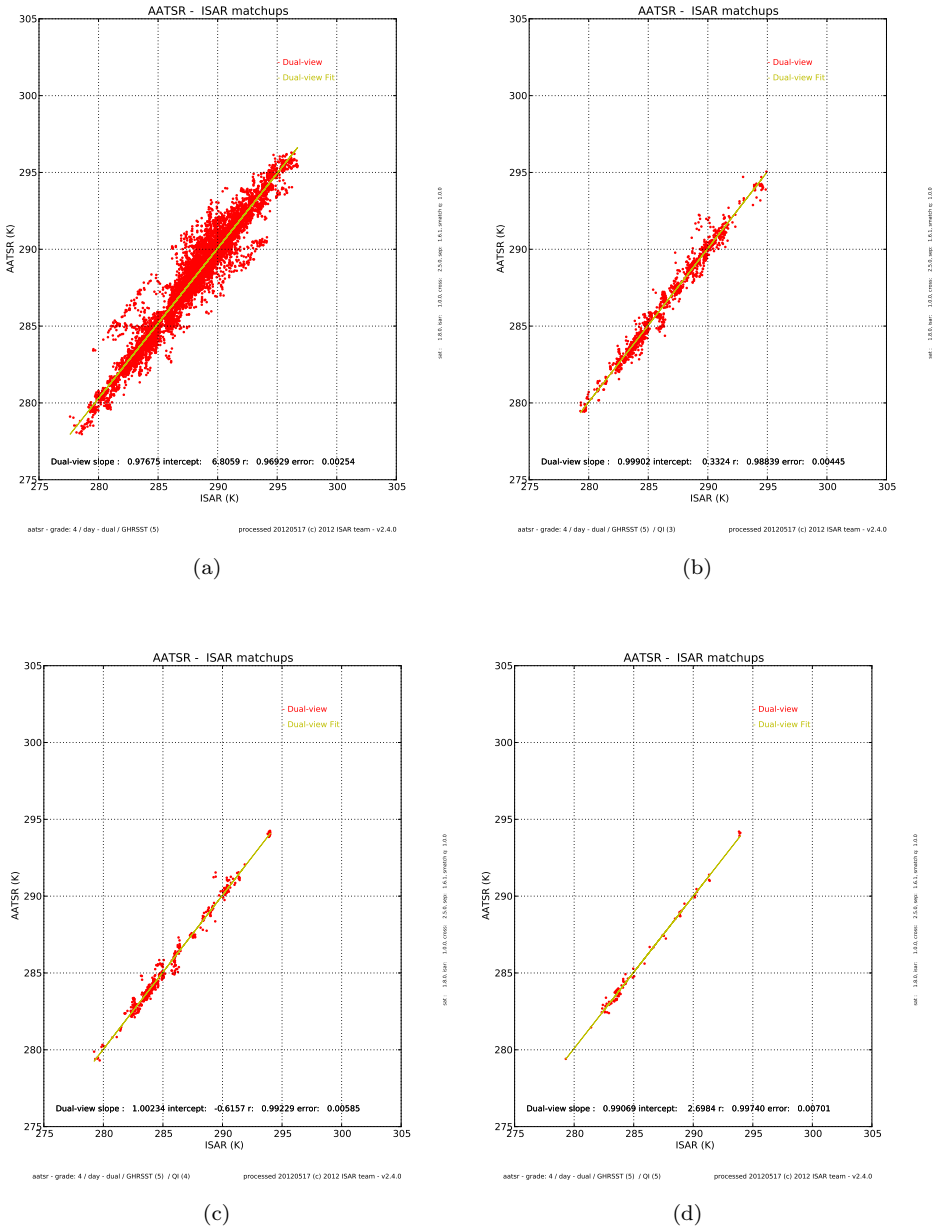


FIGURE B.23: QI scatter plot for CV 5 and normal statistics for dual view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

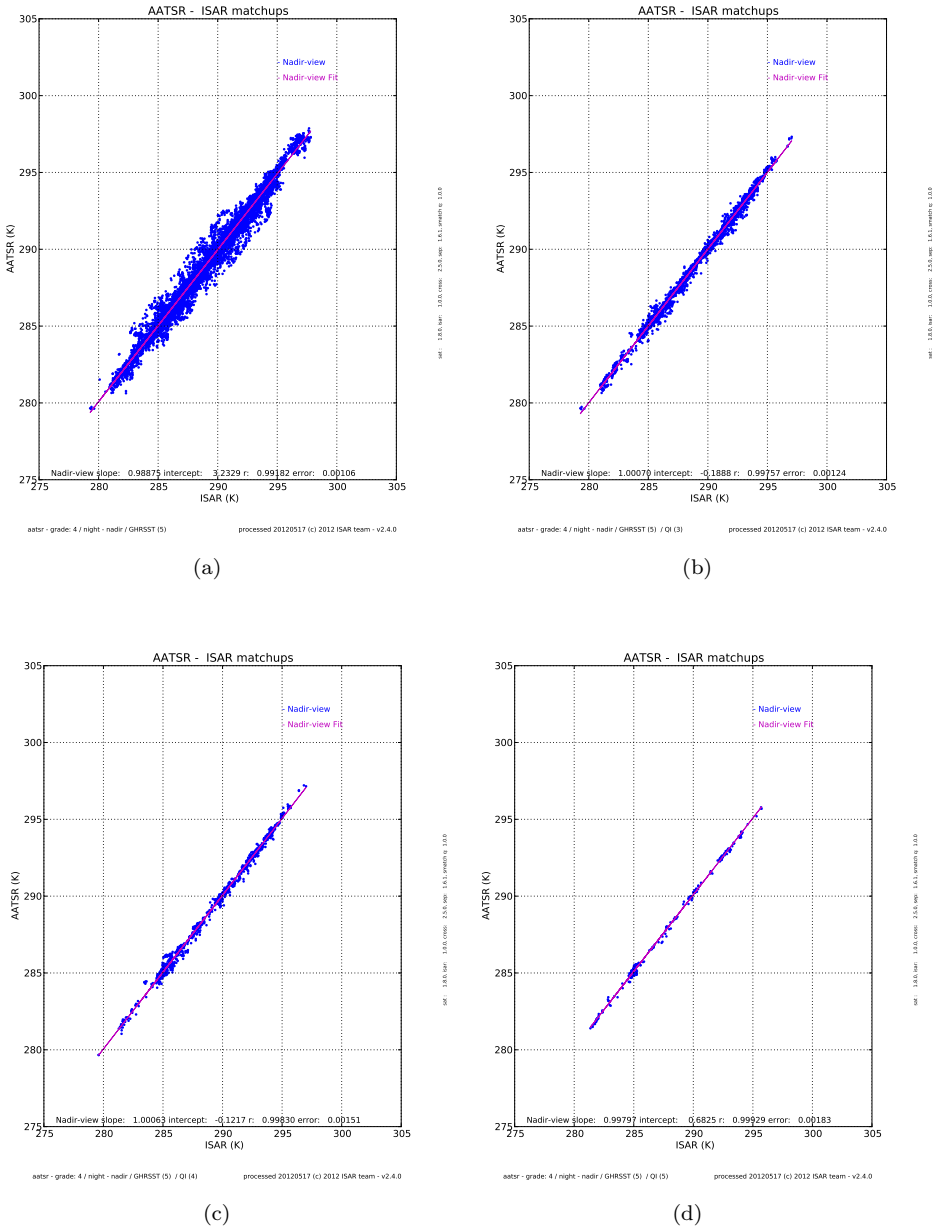
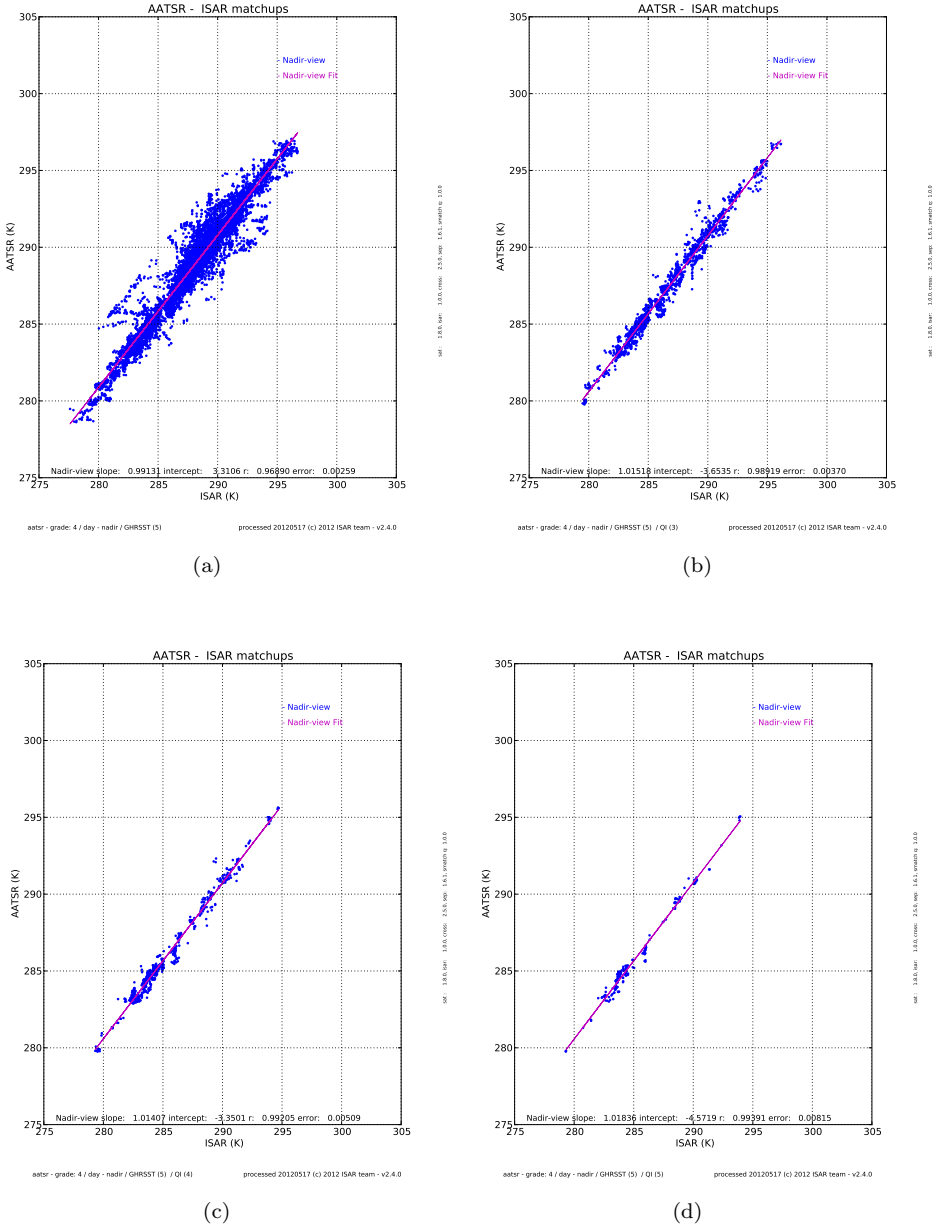


FIGURE B.24: QI scatter plot for CV 5 and normal statistics for nadir view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;



(a) (b)

(c) (d)

FIGURE B.25: QI scatter plot for CV 5 and normal statistics for nadir view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

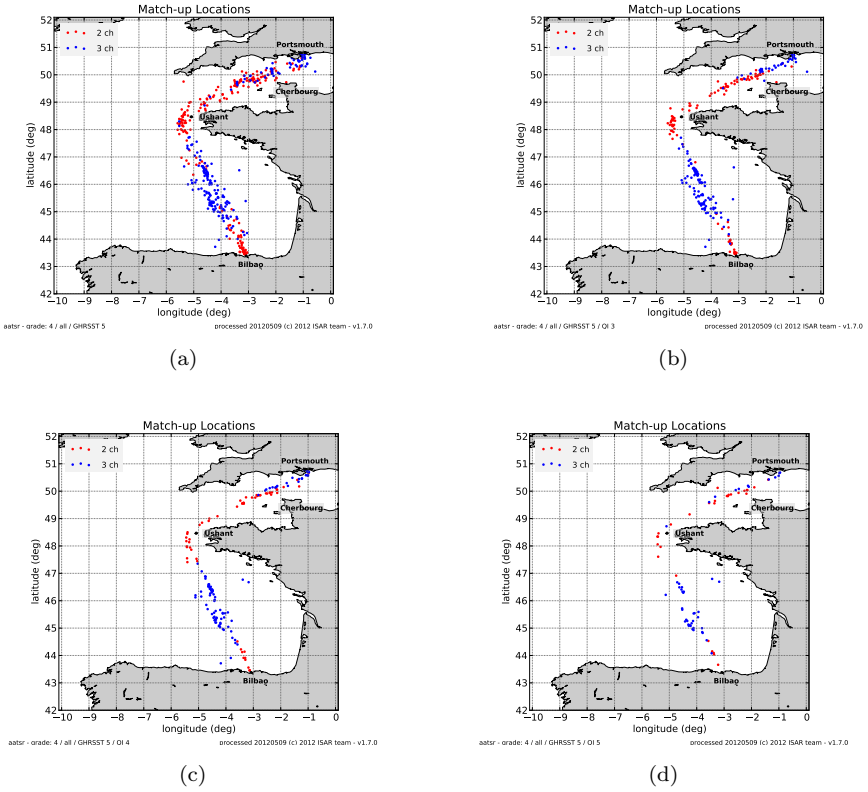
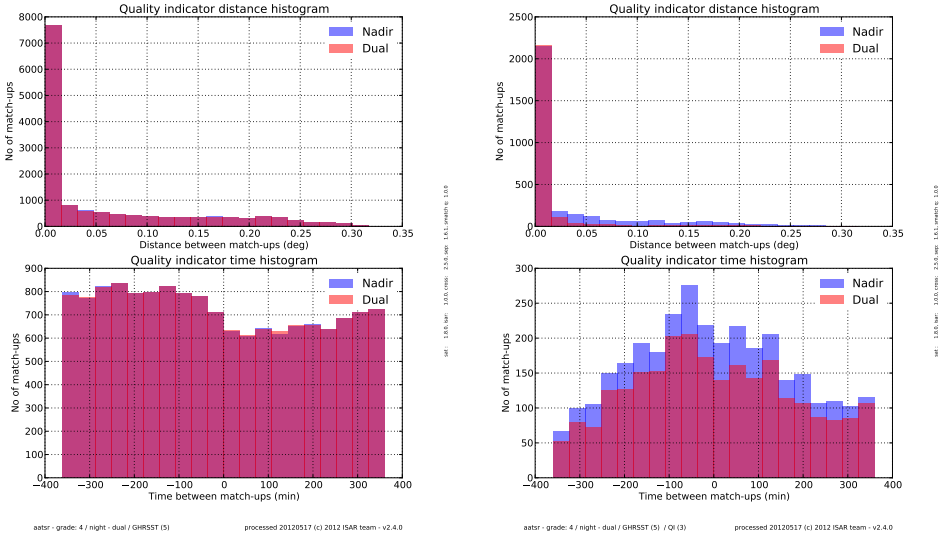
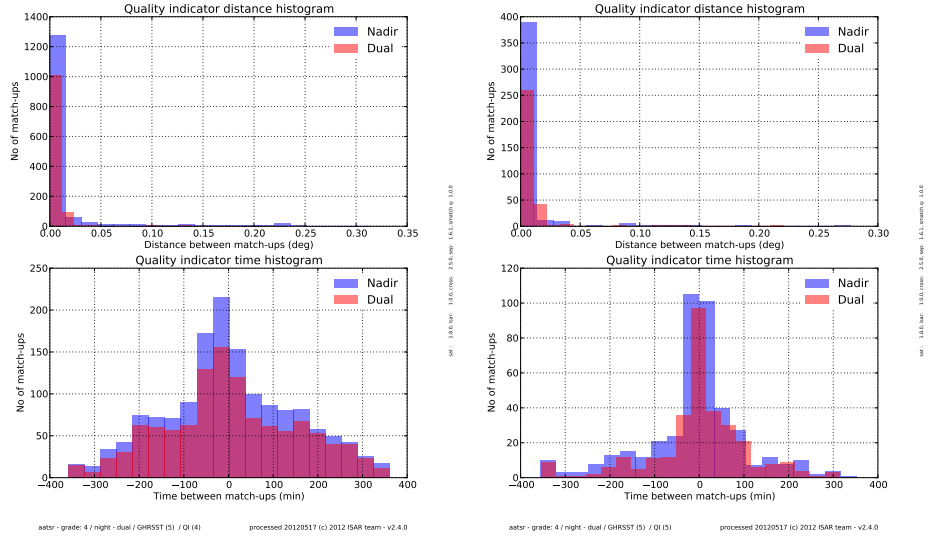


FIGURE B.26: QI location plot for CV 5 and normal statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; (c) QI 4 location plot; (d) QI 5 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue.



(a)

(b)



(c)

(d)

FIGURE B.27: QI time and distance plot for CV 5 and normal statistics for for night time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, **??** QI 5 time and distance plot; Dual view $match-up$ data are coloured red and nadir view time $match-up$ data are coloured blue.

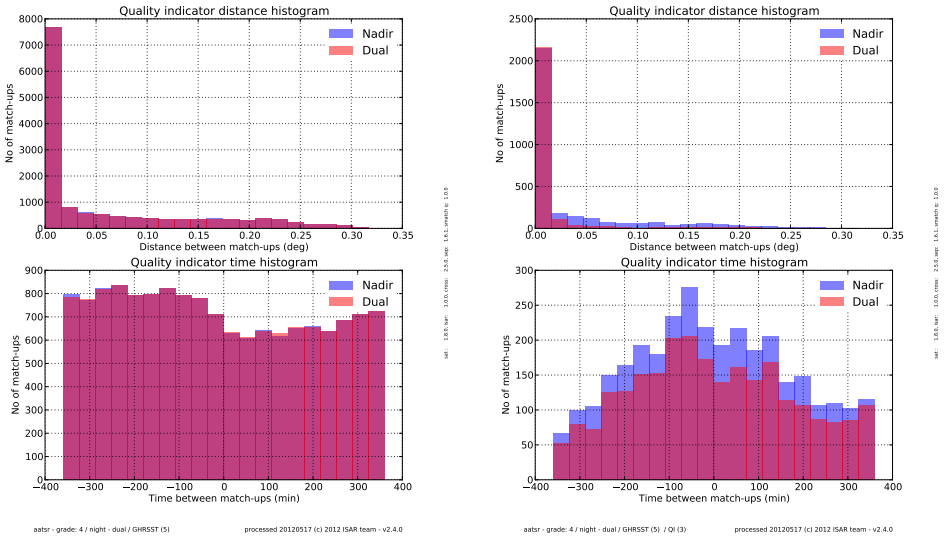


FIGURE B.28: QI time and distance plot for CV 5 and normal statistics for for day time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match-up$ data are coloured red and nadir view time $match-up$ data are coloured blue

B.2.5 AATSR N1 data with robust statistics

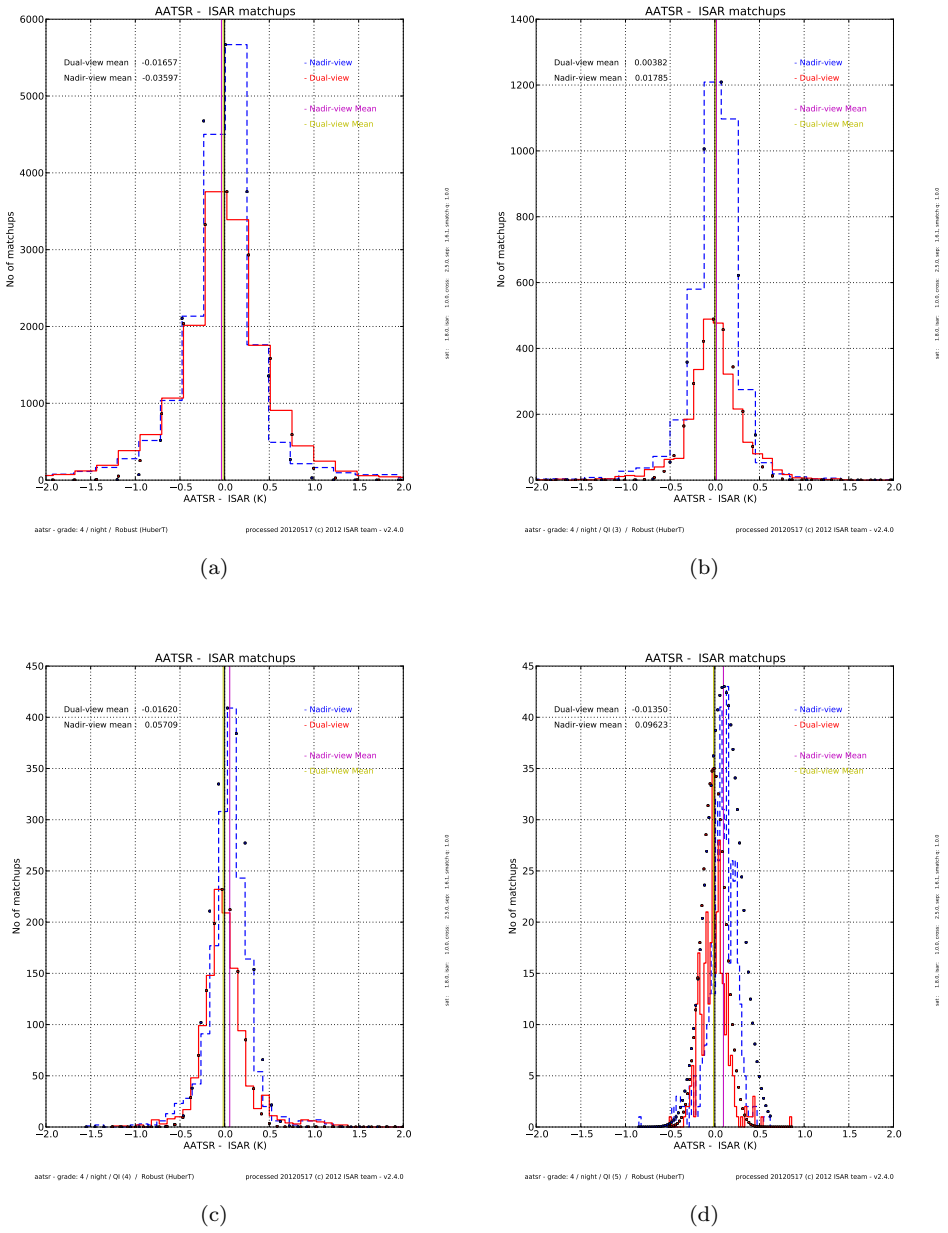


FIGURE B.29: QI histograms for N1 and robust statistics for night time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

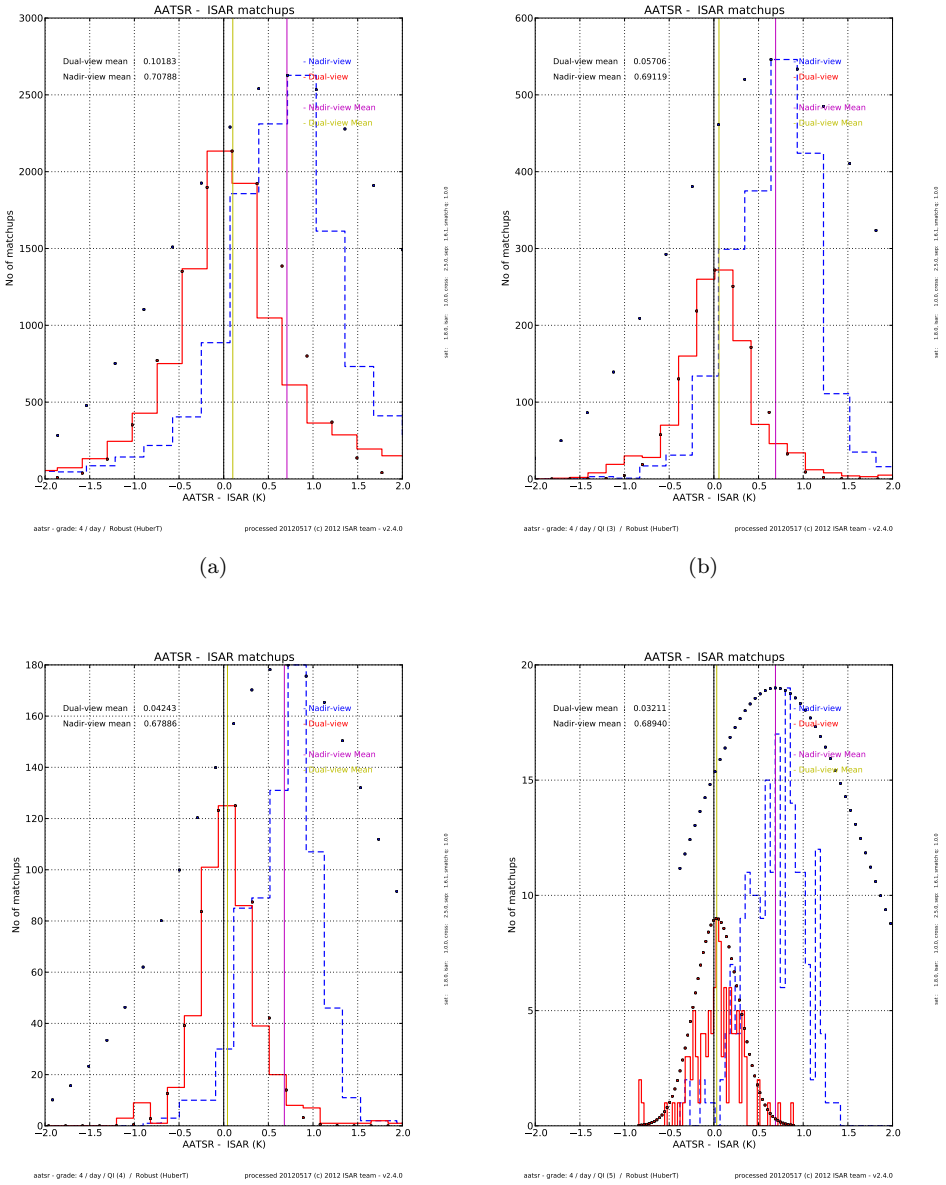


FIGURE B.30: QI histograms for and robust statistics for day time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; and, (d) QI 5 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like.

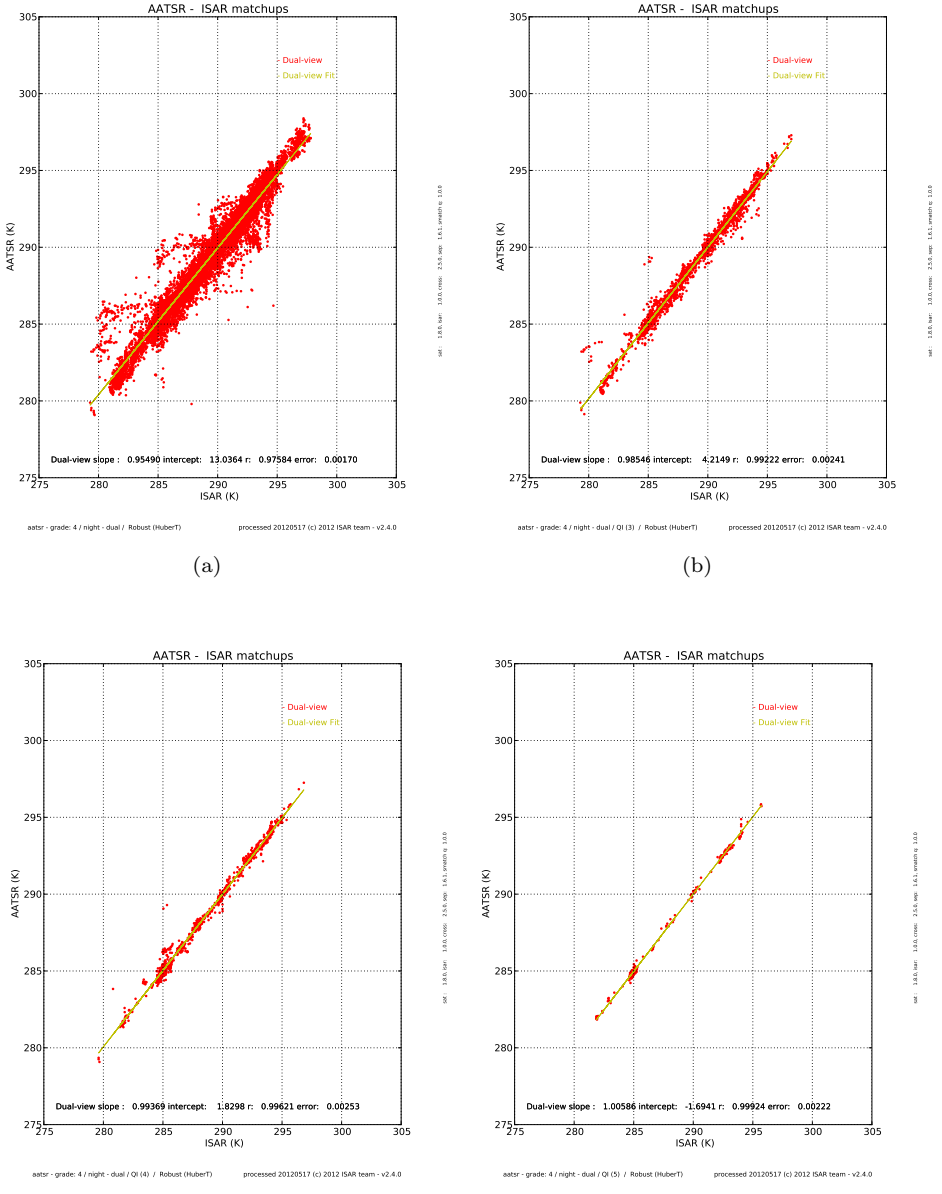


FIGURE B.31: QI scatter plot for N1 and robust statistics for dual view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

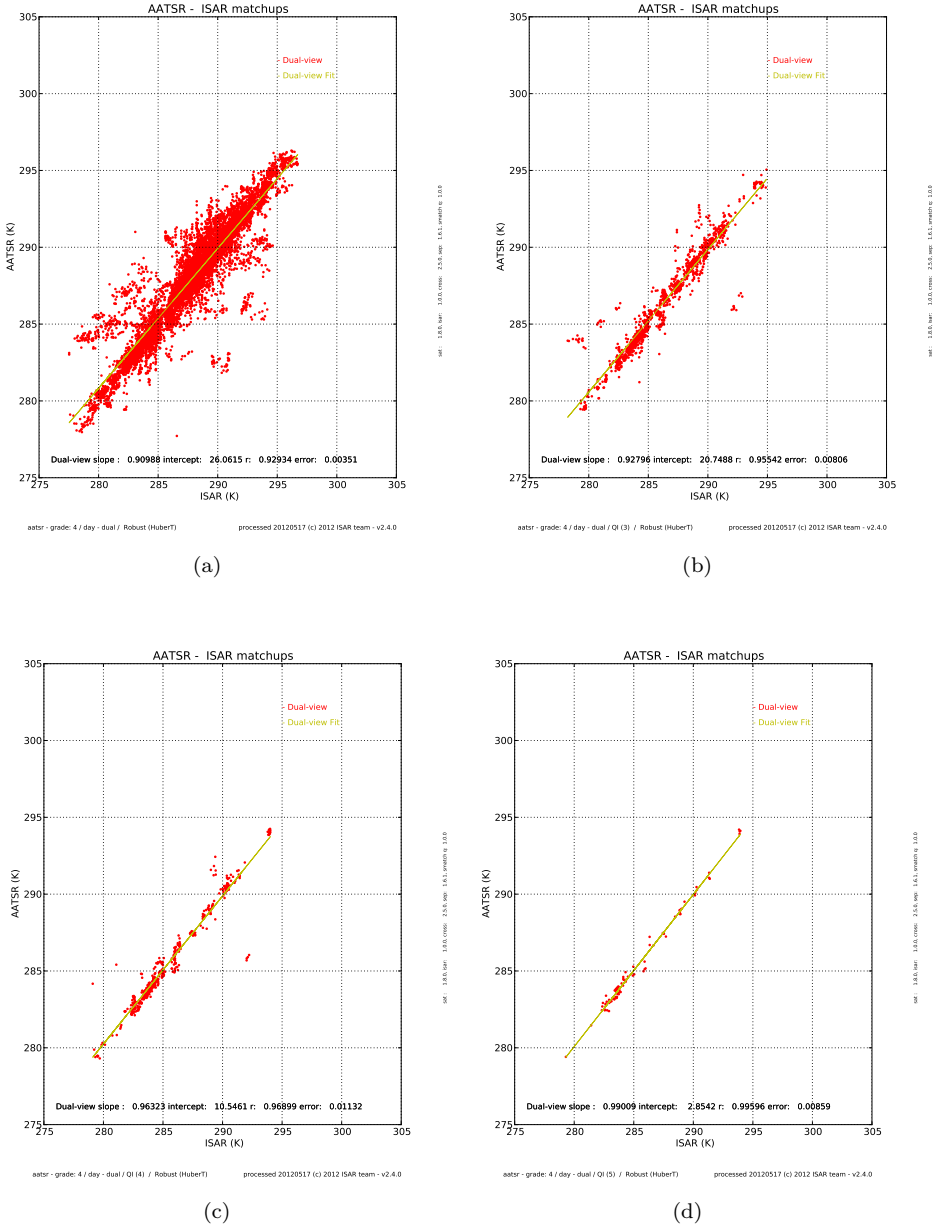


FIGURE B.32: QI scatter plot for N1 and robust statistics for dual view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

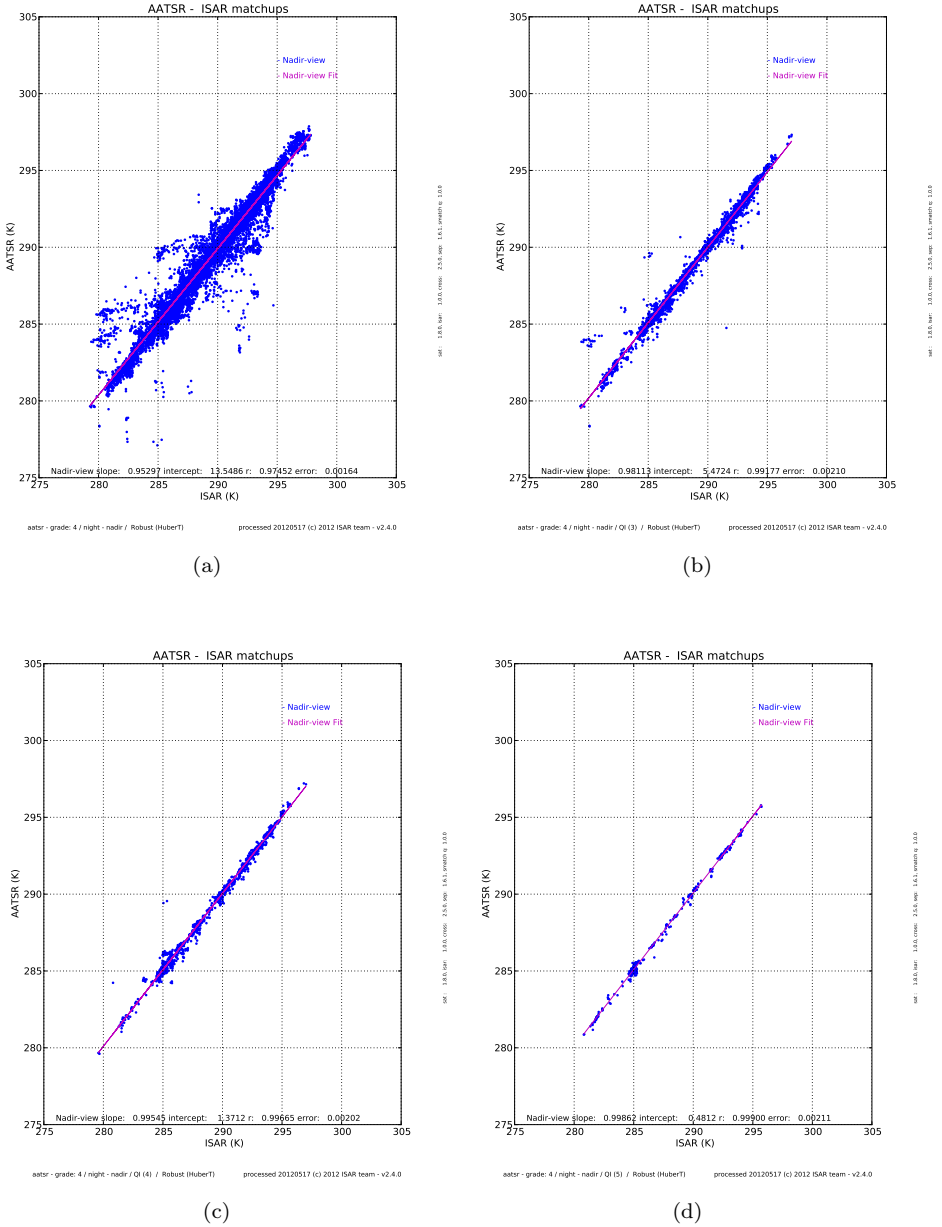


FIGURE B.33: QI scatter plot for N1 and robust statistics for nadir view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

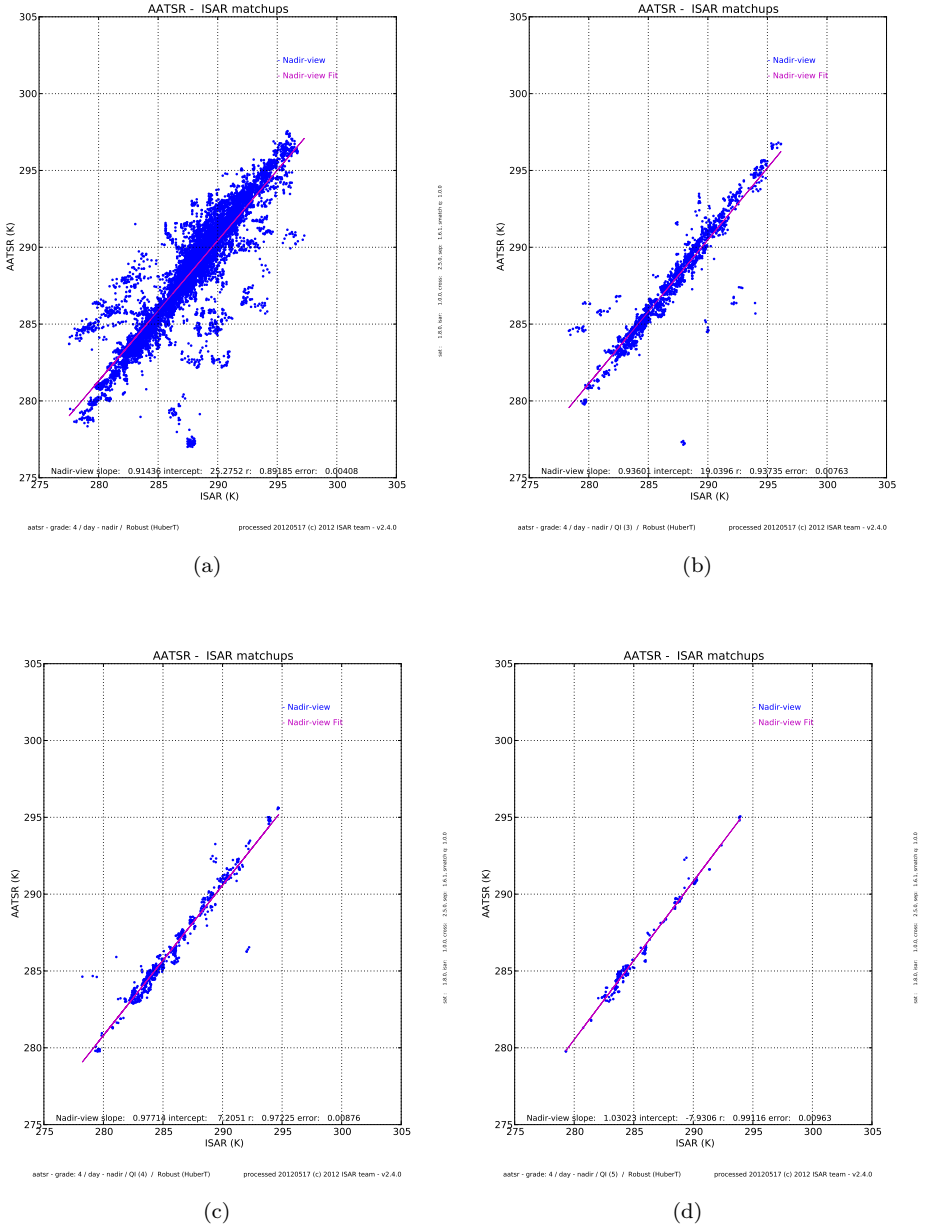


FIGURE B.34: QI scatter plot for N1 and robust statistics for nadir view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, (d) QI 5 scatter plot;

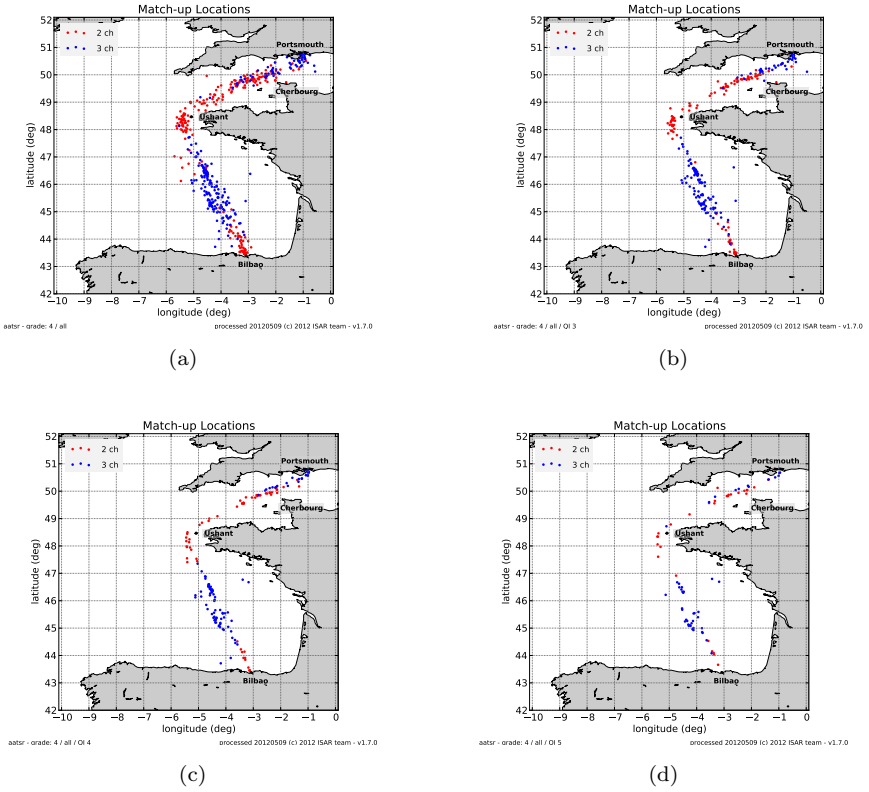
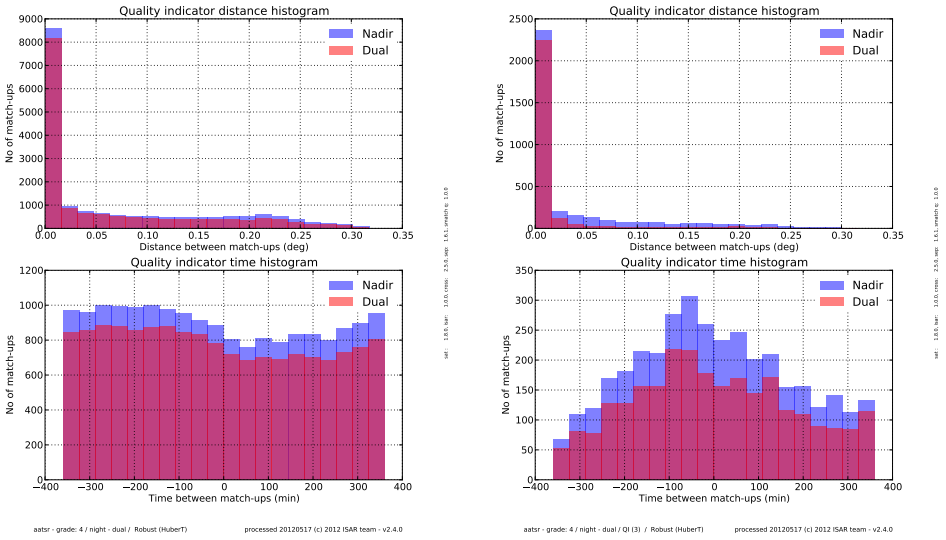
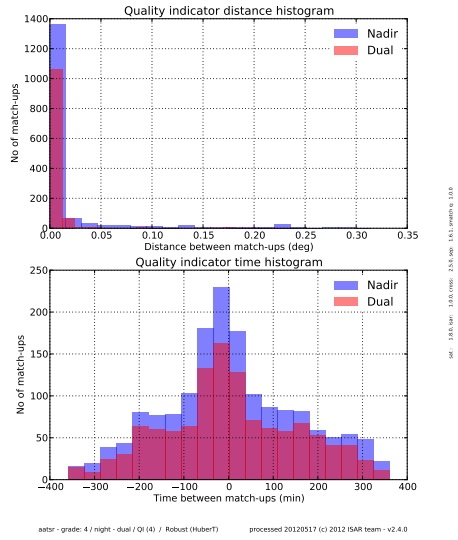


FIGURE B.35: QI location plot for N1 and robust statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; (c) QI 4 location plot; and, (d) QI 5 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue

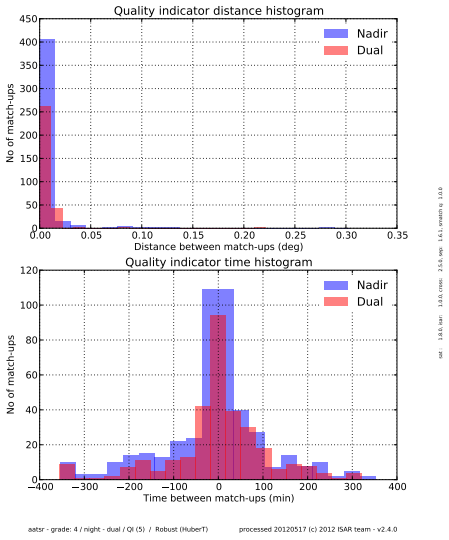


(a)

(b)

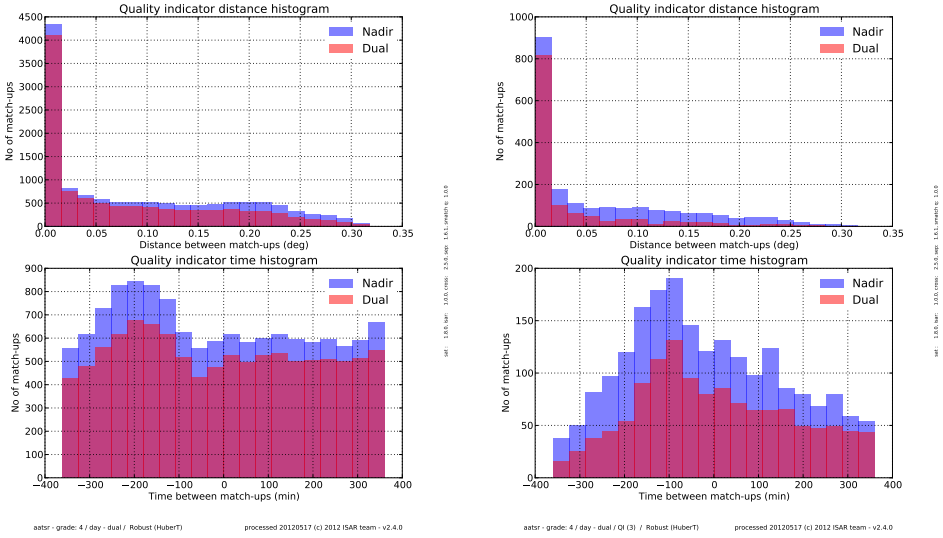


(c)



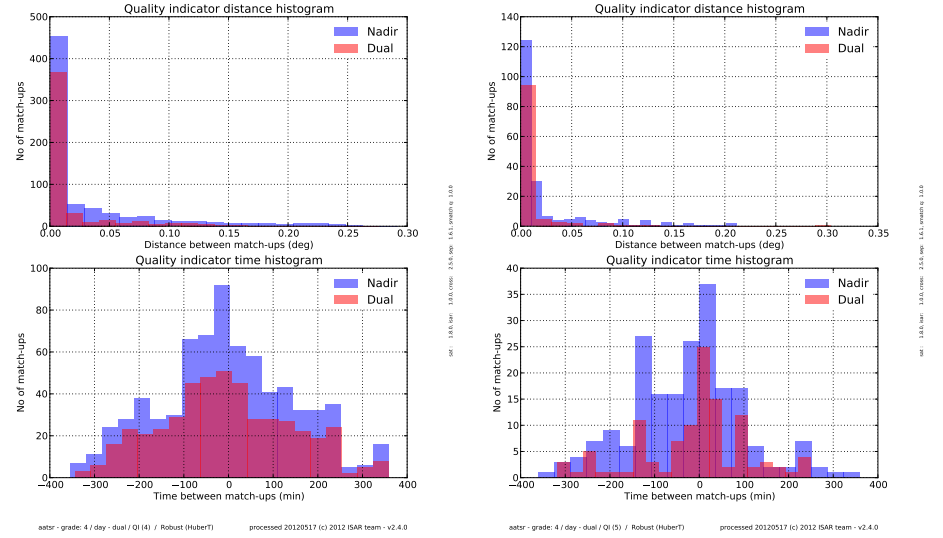
(d)

FIGURE B.36: QI time and distance plot for N1 and robust statistics for for night time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue



(a)

(b)

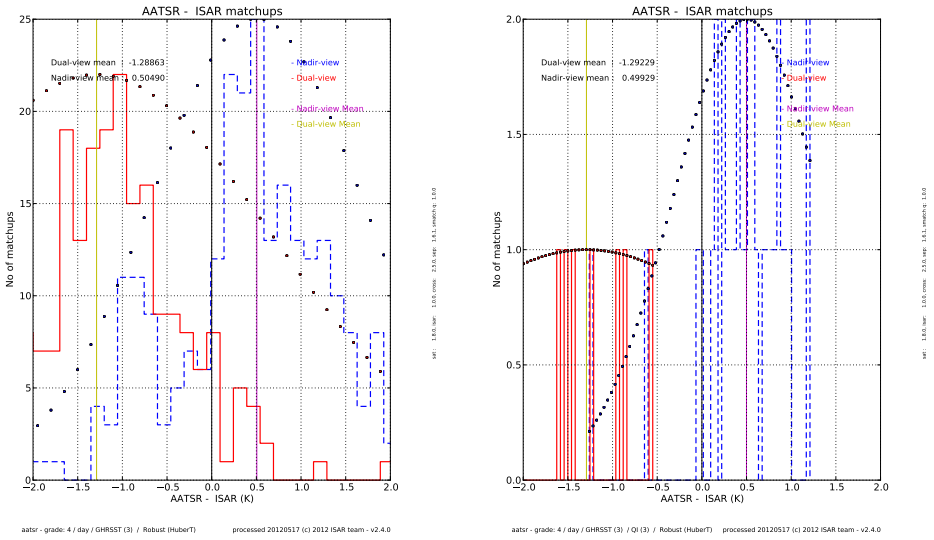


(c)

(d)

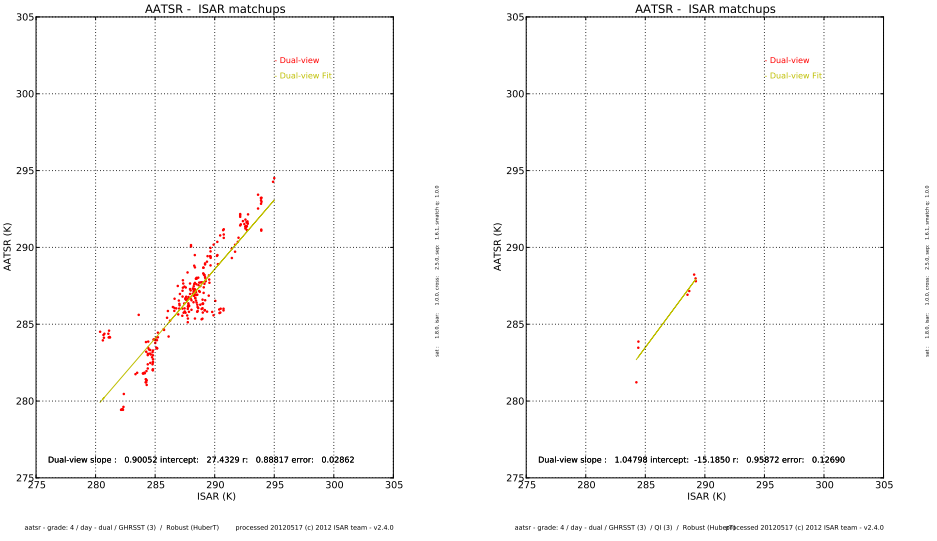
FIGURE B.37: QI time and distance plot for N1 and robust statistics for for day time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; and, (d) QI 5 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue

B.2.6 AATSR CV 3 data with robust statistics



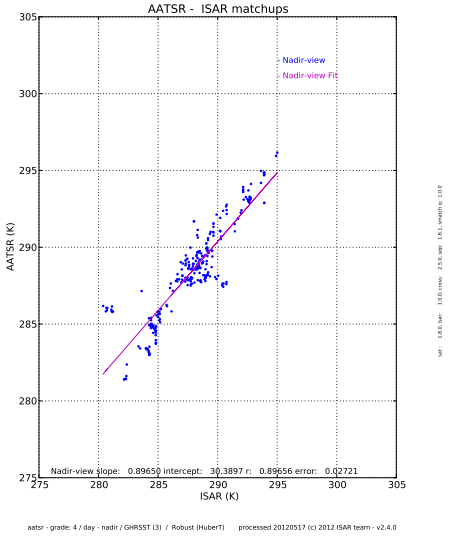
(a) (b)

FIGURE B.38: QI histograms for CV 3 and robust statistics for day time data: (a) QI 0 histogram; (a) QI 3 histogram; The red histogram shows dual view $match - up$ data, the dashed blue histogram shows nadir view only $match - up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like. No histograms for QI 4 and 5 were produced as there was no data for QI 4 and 5.

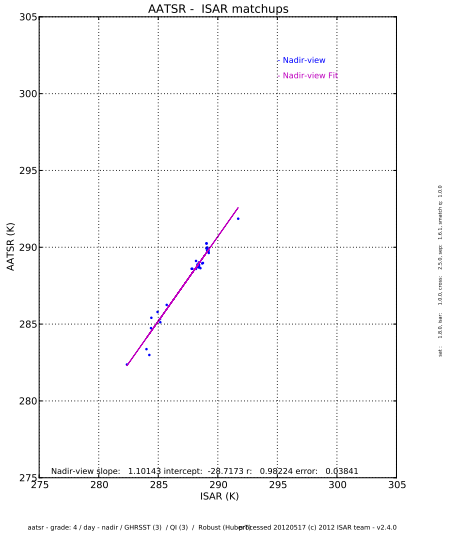


(a) (b)

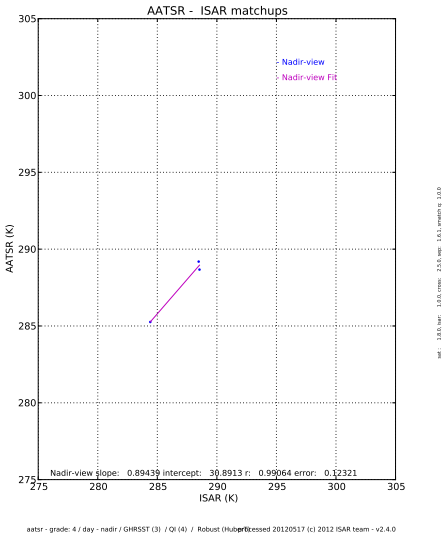
FIGURE B.39: QI scatter plot for CV 3 and robust statistics for dual view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; No QI 4 and 5 data is available.



(a)

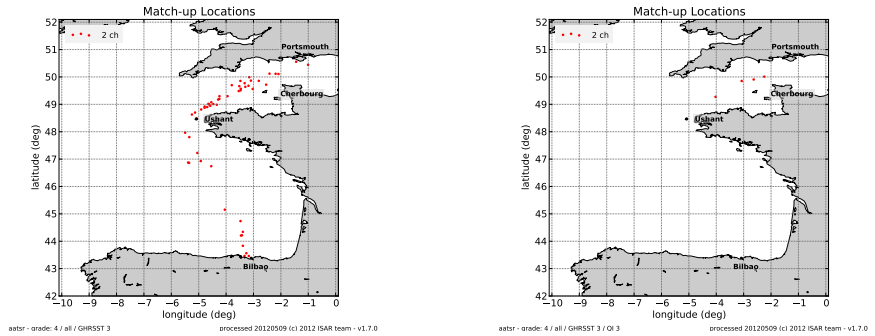


(b)



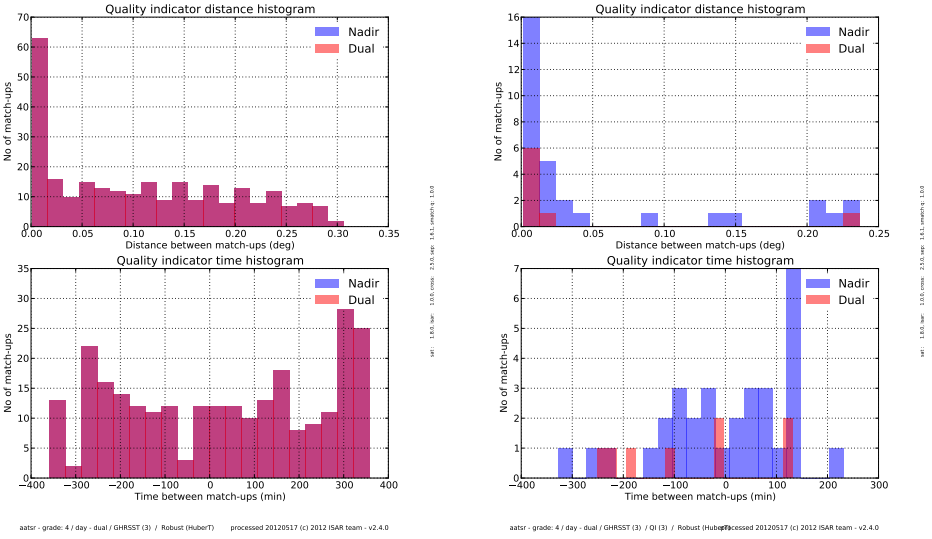
(c)

FIGURE B.40: QI scatter plot for CV 3 and robust statistics for nadir view day time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot; and, No QI 5 data was produced.



(a) (b)

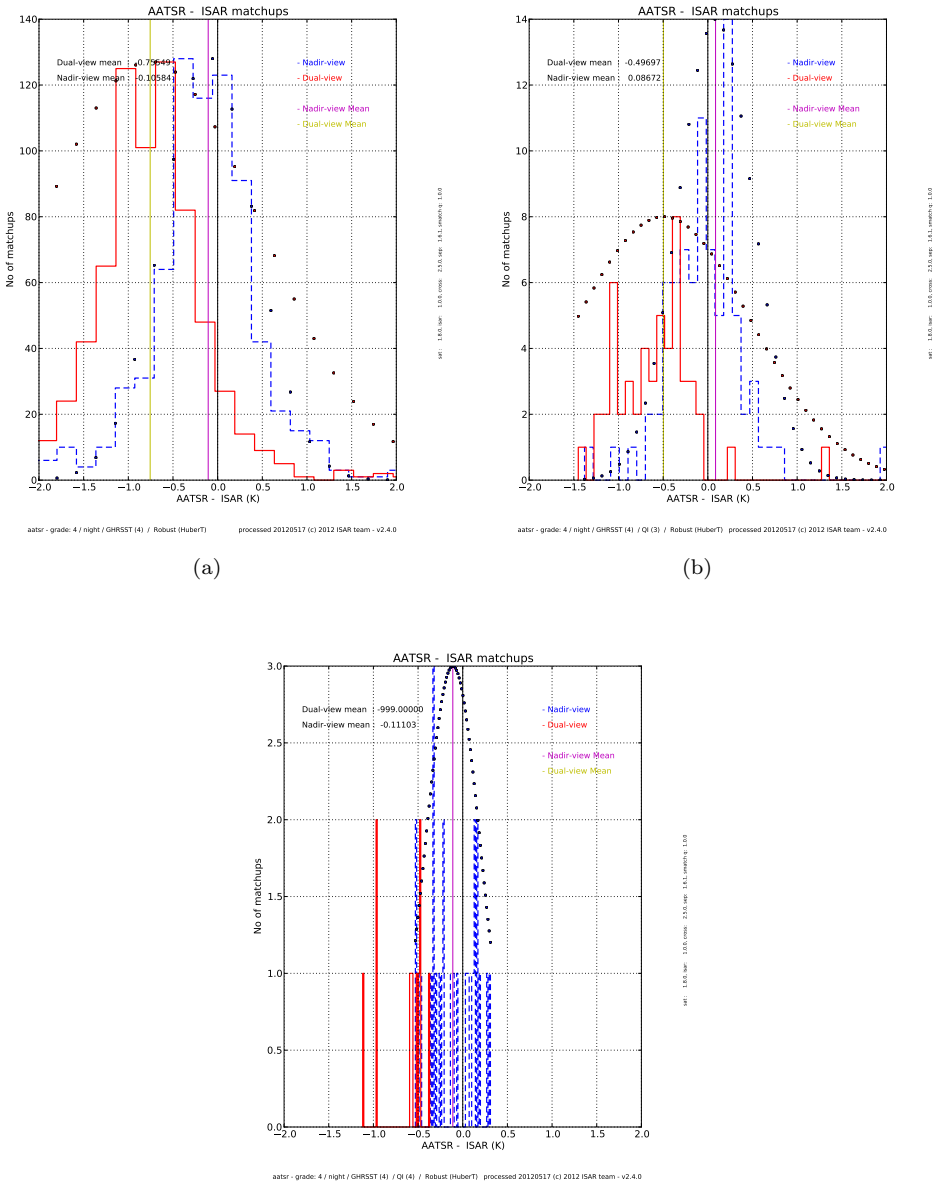
FIGURE B.41: QI location plot for CV 3 and robust statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; Day time $match-up$ data are coloured red and night time $match-up$ data are coloured blue. No QI 4 or 5 data was produced.



(a) (b)

FIGURE B.42: QI time and distance plot for CV 3 and robust statistics for for day time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; Dual view $match-up$ data are coloured red and nadir view time $match-up$ data are coloured blue. No QI 4 and 5 data available.

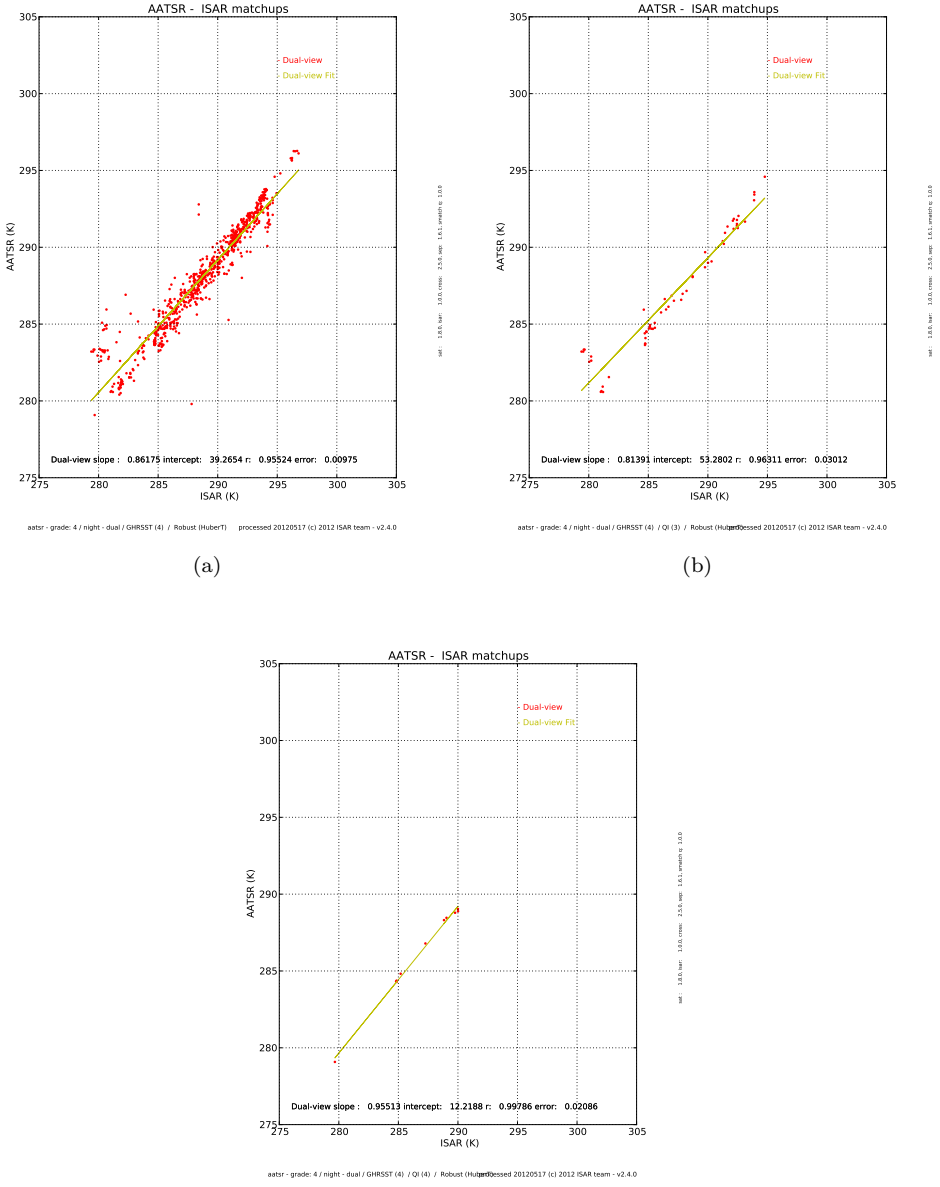
B.2.7 AATSR CV 4 data with robust statistics



(a) (b)

(c)

FIGURE B.43: QI histograms for CV 4 and robust statistics for night time data: (a) QI 0 histogram; (b) QI 3 histogram; (c) QI 4 histogram; The red histogram shows dual view $match-up$ data, the dashed blue histogram shows nadir view only $match-up$ data, the yellow vertical line shows the dual view mean and the magenta vertical line shows the nadir view only mean. The dots on the histogram show what a Gaussian distribution with the same mean and standard deviation would look like. No QI 5 data was produced

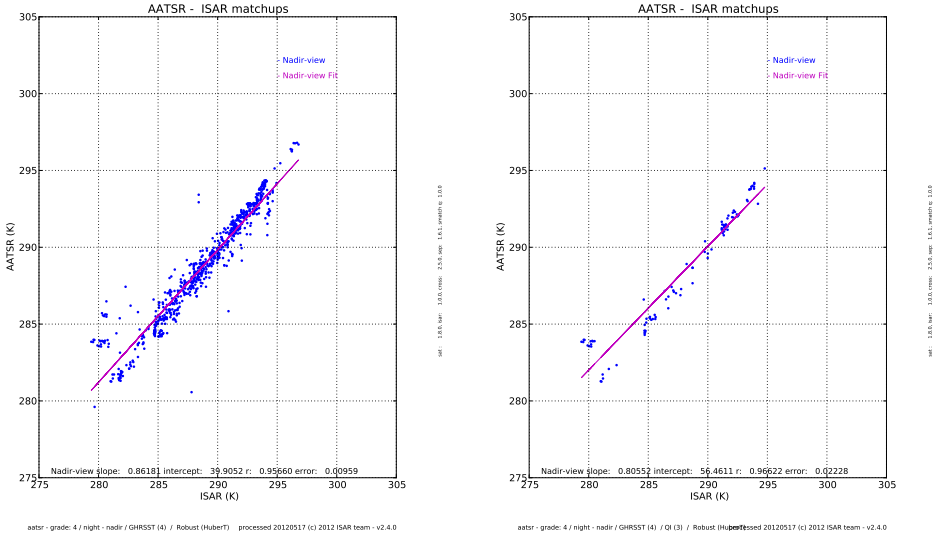


(a)

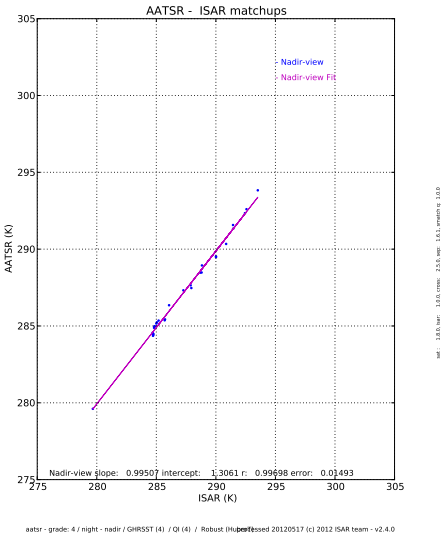
(b)

(c)

FIGURE B.44: QI scatter plot for CV 4 and robust statistics for dual view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot;No QI 5 data were produced.

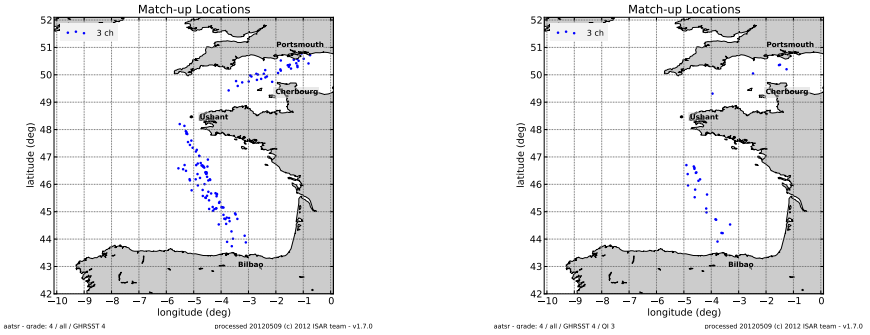


(a) (b)



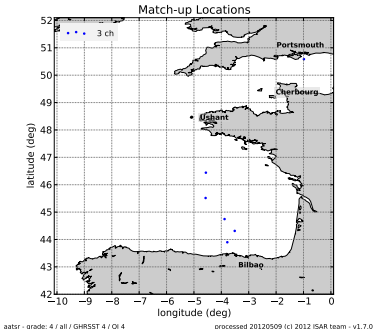
(c)

FIGURE B.45: QI scatter plot for CV 4 and robust statistics for nadir view night time data: (a) QI 0 scatter plot; (b) QI 3 scatter plot; (c) QI 4 scatter plot;No QI 5 data were produced



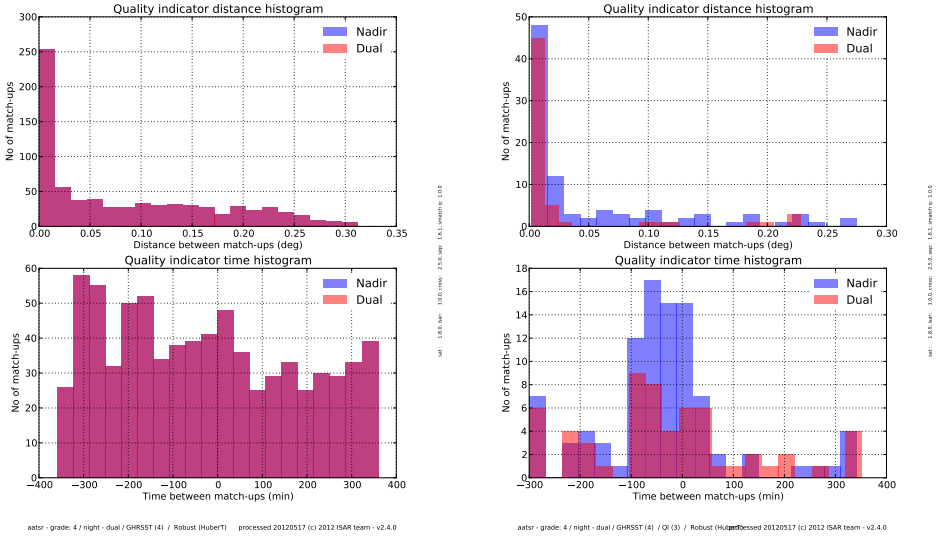
(a)

(b)



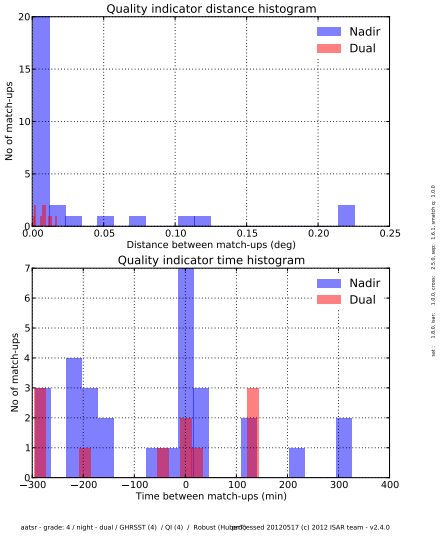
(c)

FIGURE B.46: QI location plot for CV 4 and robust statistics for for all data: (a) QI 0 location plot; (b) QI 3 location plot; (c) QI 4 location plot; Day time $match - up$ data are coloured red and night time $match - up$ data are coloured blue. No QI 5 data were produced.



(a)

(b)



(c)

FIGURE B.47: QI time and distance plot for CV 4 and robust statistics for for night time data: (a) QI 0 time and distance plot; (b) QI 3 time and distance plot; (c) QI 4 time and distance plot; Dual view $match - up$ data are coloured red and nadir view time $match - up$ data are coloured blue. No QI 5 data were produced.

Appendix C

SSES for AATSR

The GHRSST SSES methodology is described in Corlett (2007) and Corlett and Poulter (2008). A short description of the main points is given here. The SSES were developed as a way of stratifying satellite *SST* data for the GHRSST L2p data format. Table C.1 shows the five CV and their meaning. Note that for AATSR only CV levels 0, 3,4 and 5 exist. For this project the starting point for AATSR data was ESA AATSR N1 data, which is equivalent to GHRSST L2p CV 0 data. The data for CV 3, 4 and 5 was produced by implementing the scheme as described by Corlett (2007) into the ISAR project AATSR *SST* processor. This was done in order to have a greater flexibility compared the L2p data and also to be able to generate nadir only data when no dual view data is available, which is missing in the L2p data record.

Table C.1: Confidence values for L2p *SST* values and their intended meaning

CV	What it Represents
0	No data (unprocessed, land etc.)
1	No SST data (cloud etc.)
2	Bad data
3	Marginal quality
4	Good quality data
5	Excellent quality data

The values of the thresholds for AATSR for table C.3 are:

TABLE C.2: The SSES limits for AATSR

Item	Value	Unit
TU2	+0.04	K
TL2	-1.53	K
TU3	+0.51	K
TL3	-0.51	K

TABLE C.3: The SSES values for AATSR. Values for TU2, TL2, TU3 and TL3 can be found in table C.2. D - N is the difference between the dual view and the nadir view temperature measurement.

Stratification Case	**AATSR channels**	**Summary of criteria**	**Wind**	**Bias**	**Standard Deviation**	**confidence value**
Case 1	2	TL2 < D-N < TU2	$< 6ms^{-1}$	+0.20	0.33	5
Case 2	2	TL2 < D-N < TU2	$> 6ms^{-1}$	+0.20	0.33	5
Case 3	2	D-N < TL2	$< 6ms^{-1}$	-0.41	0.71	3
Case 4	2	D-N < TL2	$> 6ms^{-1}$	-0.41	0.71	3
Case 5	2	D-N > TU2	$< 6ms^{-1}$	+0.71	0.64	3
Case 6	2	D-N > TU2	$> 6ms^{-1}$	+0.71	0.64	3
Case 7	3	TL3 < D-N < TU3	$< 6ms^{-1}$	+0.11	0.32	5
Case 8	3	TL3 < D-N < TU3	$> 6ms^{-1}$	+0.11	0.32	5
Case 9	3	D-N < TL3	$< 6ms^{-1}$	-0.65	0.49	4
Case 10	3	D-N < TL3	$> 6ms^{-1}$	-0.65	0.49	4
Case 11	3	D-N > TU3	$< 6ms^{-1}$	+0.69	0.32	4
Case 12	3	D-N > TU3	$> 6ms^{-1}$	+0.69	0.32	4

Appendix D

Validation paper

Remote Sensing of Environment 116 (2012) 17–31

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Long-term validation of AATSR SST data products using shipborne radiometry in the Bay of Biscay and English Channel

Werenfrid Wimmer [a], Ian S. Robinson [a,*], Craig J. Donlon [b]

[a] School of Ocean and Earth Science, National Oceanography Centre, University of Southampton Waterfront Campus, European Way, Southampton SO14 3ZH, United Kingdom
[b] European Space Agency ESTEC Facility, Keplerlaan 1, 2201 AZ Noordwijk, Netherlands

ARTICLE INFO

Article history:
Received 22 September 2009
Received in revised form 11 March 2011
Accepted 12 March 2011
Available online 29 July 2011

Keywords:
AATSR
Sea surface temperature
Validation
Shipborne radiometry
Bay of Biscay
English Channel

ABSTRACT

Radiometric measurements of the sea surface temperature (SST) made by the infrared SST autonomous radiometer (ISAR) deployed routinely between 2004 and 2009 from a passenger vessel traversing the English Channel and Bay of Biscay are used to validate satellite SST data products produced using the Advanced along-track scanning radiometer (AATSR) flown on Envisat. More than 1500 independent pairings between an ISAR record and an AATSR pixel, coincident within specified space–time matching windows, are analysed. These confirm good agreement between the in situ and the satellite derived SST estimates, based on the dual view AATSR algorithms, with a bias of less than 0.1 K which is the accuracy limit of the ISAR. The standard deviation of the comparisons depends on the coincidence criteria: for a match-up window of 1 km and 2 h it is around 0.3 K for the three channel (night only) algorithm and 0.4 K for the 2 channel algorithm. Separate validation statistics are produced for the periods before and after 7 Dec 2005 when a change was made to the AATSR algorithms. It is shown that the error distribution was narrowed by introducing the new algorithm and further narrowed by using only the AATSR data that have the highest Confidence Value. This is the first systematic use of autonomous underway shipboard radiometry on a vessel of opportunity for validating satellite data. The methodology is carefully assessed and shown to provide an effective and reliable means of confirming the high quality and stability of the SST data products from AATSR.

1. Introduction

This paper presents a consistent approach to validating measurements of sea surface temperature (SST) from satellites, based on the acquisition of matching data using a purpose-built infrared radiometer that operates autonomously from ships of opportunity. Specifically it reports the first systematic mid-latitude validation of the Advanced Along-Track Scanning Radiometer (AATSR) based on more than five years of regular observations acquired in the Bay of Biscay and the English Channel. More generally it addresses the principles of constructing match-up datasets between satellite observations and ship-based measurements along regular transects of ocean-going vessels.

The AATSR is a visible and infra-red sensor deployed since 2002 on the European Space Agency (ESA) Envisat satellite, with the primary objective of accurately measuring SST. Like its predecessors the Along-Track Scanning Radiometer (ATSR) (Edwards et al., 1990) and ATSR-2, this radiometer scans conically, making two independent observations of each part of the sea surface within the 512 km swath, through different atmospheric path lengths. With two infrared channels in the 10.0–12.5 μm atmospheric window and another centred on 3.7 μm there are six independent measurements of brightness temperature for each pixel viewed at night and four in daylight when the 3.7 μm channel cannot be used due to reflected solar contamination. This provides a unique opportunity for improved atmospheric correction (Závody et al., 1995), which is robustly sensitive not only to the effect of water vapour but also to stratospheric aerosols (Merchant et al., 1999; Merchant & Harris, 1999). We shall refer in this paper to the dual-view SST data products as Dual-3 when three spectral channels are used (night-only) and Dual-2 when two channels are used. SST products are also derived using the nadir view only, and these products are referred to as Nadir-2 and Nadir-3.

The dual-view SST data products (Dual-2 and Dual-3) from the AATSR are increasingly being recognised as providing greater stability and accuracy than most other satellite SST products although their daily coverage is restricted by the narrower swath. Facilitated by the common data format and standards pioneered by the GODAE high resolution SST pilot project, GHRSST-PP (Donlon et al., 2007), AATSR data are now used to provide a stable reference against which to adjust the bias of SST products from other satellite sensors in operational SST analyses that blend inputs from several satellites (Robinson et al., in press).

If AATSR data are to be effective in this role of enhancing the accuracy of SST products blended from several sources, their own

* Corresponding author.
E-mail address: isr@noc.soton.ac.uk (I.S. Robinson).

0034-4257/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2011.03.022

accuracy statistics need to be well characterised. Unlike SST retrieval algorithms for most other infra-red and microwave sensors on satellites, the (A)ATSR approach does not make any use of in situ temperature measurements to calibrate or tune the SST retrieval algorithm or for applying a bias correction. Instead the algorithm is derived from first principles based on radiation transfer physics and delivers an estimate of the skin temperature of the sea. In principle this independence from in situ observations should improve the inherent accuracy of the retrieved SST. It eliminates potentially large errors arising from unmeasured differences between the skin temperature observed by a satellite radiometer and the temperature recorded by a contact thermometer just beneath the surface, as discussed by Donlon et al. (2002) and Robinson and Donlon (2003). However, in practise the accuracy of AATSR data must be validated against independent measurements of SST if users are to have confidence in them as a reference standard for other SST datasets. Hence it is essential that validation of AATSR SST data products be performed regularly through the lifetime of the sensor.

Much previous validation work for SST retrievals from ATSR, ATSR-2 and AATSR has had to depend on contact thermometers on buoys or research vessels (O'Carroll et al., 2006). Although this approach can monitor the stability and uniformity of the SST data product, it does not eliminate uncertainties associated with the variable thermal structure of the upper few metres of the ocean. Such uncertainties make it difficult to determine the absolute error statistics of AATSR, and especially its bias and standard deviation to within 0.1 K which is a desirable goal if the use of AATSR as a reference standard for other SST products is to be justified, or if data from the ATSR series are to provide an independent climate record of SST (Merchant et al., 2008). Shipborne infra-red radiometry offers the potential to obtain a matched set of in situ and coincident satellite data that can be used for validating the AATSR with reduced uncertainty since issues of the different depth sampled by conventional in situ SST sensors are completely eliminated. For this reason the Infrared Sea surface temperature Autonomous Radiometer (ISAR) was developed for use on ships of opportunity (Donlon et al., 2008).

The primary purpose of this paper is to present the results of the validation over several years of AATSR SST products using coincident SST skin temperature observations made by the AATSR and by the ISAR deployed on the *Pride of Bilbao*, a car ferry that makes a return trip between Portsmouth and Bilbo twice a week. Whilst some preliminary AATSR validation results were presented by Donlon et al. (2008), in this paper we are able to make a far more definitive assessment of the AATSR performance and its stability over several years, based on a large number of samples. The secondary aim is to assess the effectiveness and value of using a shipborne radiometer for satellite SST validation and to review whether it is desirable to widen the geographical extent of this approach.

Previous attempts to validate the SST products from ATSR and ATSR-2 using earlier shipborne radiometers showed the promise of this approach (Barton et al., 1995; Donlon & Robinson, 1998; Parkes et al., 2000) but were restricted mainly by the small number of matches that could be achieved, partly because the cost of staff required to attend non-autonomous radiometers severely limited the number of deployments. A more substantial regional validation of the AATSR in the Caribbean (Noyes et al., 2006) used shipborne radiometric measurements during 2003 from the Marine-Atmosphere Emitted Radiance Interferometer (M-AERI), an interferometric infrared radiometer mounted on the cruise vessel *Explorer of the Seas*. That work acquired 99 match-ups between the in situ and satellite data, 74 at night for the Dual-3 product and 25 in the day for Dual-2. A number of occasions were detected when the presence of Saharan dust in the atmosphere produced a warm bias in the dual view SST products. When those events were eliminated, the differences between the satellite and the ship radiometer were found to be 0.02 K ± 0.25 K for Dual-3, and −0.05 K ± 0.26 K for Dual-2.

A summary of all the available validation observations covering the first two years of AATSR operation from 2002 to 2004 (Corlett et al., 2006) contains a combination of buoy data and global compilations from ships of opportunity, as well as a small number of observations from a few ship radiometers, including early examples from the dataset presented here. Overall it concluded that the ATSR series of instruments continued to be the world leader in delivering accurate space-based observations of SST. The standard deviation of the satellite − in situ temperature differences was less than 0.3 K for ship radiometry and slightly greater for contact thermometry. Nonetheless the data sources, geographic distribution and temporal distribution of the validation data across that two year time span is heterogeneous, making it difficult to test for systematic spatial and/or temporal trends or patterns in the validation statistics. In the context of the ongoing global validation of AATSR to which they contribute, the five years of data presented here provide a unique opportunity to validate the longer term stability of the AATSR performance, albeit in a limited sea area. It also allows an objective evaluation of a small change to the AATSR retrieval algorithm that was made on 7 December 2005.

2. Experimental approach

The ISAR is installed on the P&O Ferry *Pride of Bilbao*, which sails regularly from Portsmouth in the UK to Bilbao in northern Spain and back, crossing the English Channel and the Bay of Biscay. The ISAR instrument is mounted on the top of the starboard bridge deck together with a set of ancillary instruments including an anemometer, short and long wave downwelling radiometers and air temperature and humidity sensors. The *Pride of Bilbao* is scheduled to be at sea most of the time with only brief turnarounds in port, thus maximising the opportunities for match-ups with AATSR, except when the vessel is laid up for refit and maintenance for a few weeks each winter. Over a regularly repeated three day cycle, the vessel traverses both a region of shelf sea (the English Channel and Celtic Sea) and the deep Atlantic Ocean (Bay of Biscay), as shown in Fig. 1.

The ISAR instrument is a single channel (waveband, 9.6–11.5 μm) scanning radiometer with two internal calibration black-bodies (Donlon et al., 2008). For this validation programme it views the sky and sea at zenith angles of 25° and 155° respectively, and the sea surface emissivity corresponding to the 25° incidence angle is taken as 0.9916 (Niclòs et al., 2005; Wu & Smith, 1997). Two radiometers are used interchangeably, being switched approximately every three months to allow inspection, servicing and replacement of any worn or optically degraded parts. The exposed optical elements of ISAR, including the scan mirror, are protected from bad weather by a shutter that closes when precipitation, spray or excessive atmospheric dust is detected by an optical rain gauge, so that ISAR can operate autonomously. In the early years of the project the ISAR was inspected in port every two weeks to ensure that it was functioning properly. In later years a subset of data including diagnostics were transmitted to NOCS every four hours by Iridium satellite communication so that the data quality could be monitored and if necessary a deployment could be ended prematurely the next time the ship berthed at Portsmouth.

Measurements of the skin SST are made by the ISAR approximately once every 140 s, following the procedure explained by Donlon et al. (2008). For the measurements reported here the internal calibration procedure was refined to allow for increased diffuse scattering by the scan mirror as it slowly degrades (see Appendix A).

The quality of the SST measurements made by ISAR is assured by an external validation procedure that is performed on the ISAR instrument in the laboratory before and after every deployment. This involves the viewing of CASOTS-2, a specially-designed high quality blackbody, over a period of several hours whilst its temperature is slowly raised. Comparison between the ISAR-derived temperature and the blackbody temperature as measured by a platinum resistance

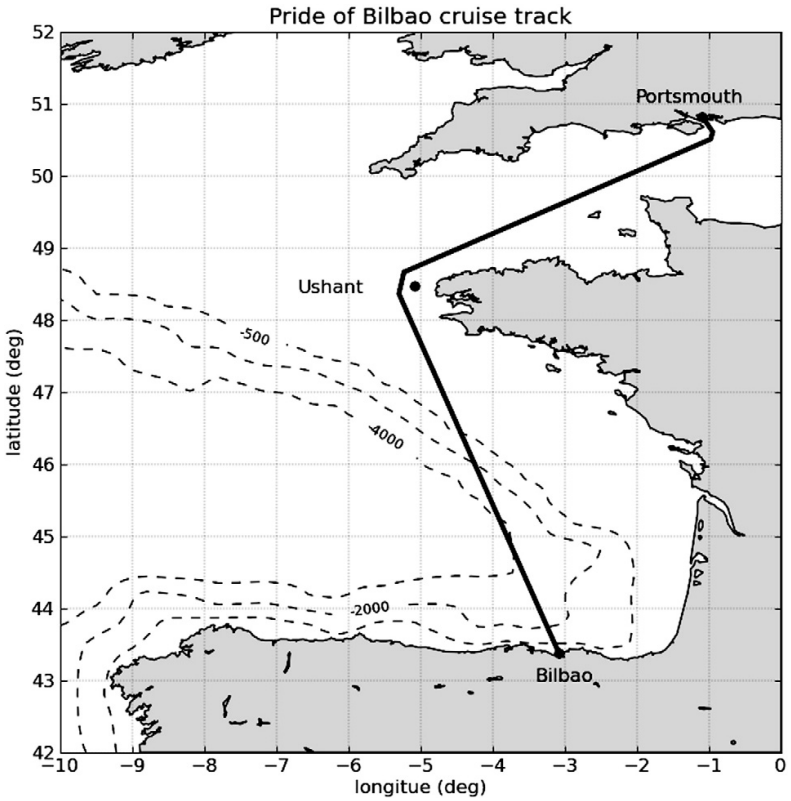


Fig. 1. Map of the study area. The solid black line shows a typical ship track and the dashed lines show the depth contours for 500, 2000 and 4000 m.

thermometer provides a characterisation of the ISAR's internal calibration process. The ISAR is not deployed unless this laboratory validation shows agreement between the ISAR and the external blackbody temperatures that is consistently better than 0.1 K. Similarly, if the post-deployment laboratory calibration does not show the same agreement, the data acquired during that deployment are not used for AATSR validation purposes unless a thorough analysis can identify and eliminate the source of the discrepancy. Typical examples of calibration results are presented by Donlon et al. (2008). On some occasions the validation is performed in a temperature controlled room to ensure that the ISAR is validated across the whole range of target temperatures encountered in the study area.

In order to tie the external validation to the same absolute radiometric temperature standards as the AATSR instrument itself, the CASOTS-2 blackbody has been periodically subjected to an international intercalibration exercise (see, for example, Rice et al., 2004) at which it is referenced to a NIST standard.

Fig. 2 sets out the record of individual ISAR deployments from March 2004 to April 2009, showing the numbered deployments (D1 etc.) and which of the two instruments was operating. Apart from a few malfunctions in the early deployments, which were rectified by component design changes following D2, and the annual 4–5 week vessel refits, one of the instruments has always been operating on the *Pride of Bilbao*. However there were no useful data returned from four aborted deployments, because of an electronics failure in D5, an instrument configuration error in D9, a mechanical failure in D15 and a power supply failure in D21. At all other times the ISAR has made a measurement of the sea temperature every 140 s if the shutter was open, or recording black body temperatures every 15 min when the shutter was closed.

Fig. 3 is a latitude-time plot of the complete five-year ISAR skin temperature record. This shows in the x axis the position (in latitude) along the cruise track and, in the y axis, the date-time. The colours represent the measured skin temperature. As expected the SST record is warmer towards the left of the figure which is the Bilbao end of the ferry's route, and cooler to the right (the Portsmouth end). The gaps in the dataset are mainly due to adverse weather, especially in the winter time, and the yearly refit which has a varying length. Over the five years the figure shows that the ISAR encountered temperatures as low as about 5 °C and as high as about 26 °C. The seasonal pattern is evident in the figure, whilst it appears that ISAR sampled the highest temperatures in 2006.

3. Match-up methodology

Having secured the operational supply of in situ skin SST along the *Pride of Bilbao* cruise track, these data were then matched to coincident AATSR observations, with which they were paired to produce a match-up dataset. The size of this dataset depends on how stringent is the definition of data coincidence in space and time. If the requirement is too severe then there will be very few matches. If it is not severe enough then a large number of the matches will pair SST measurements from slightly different places in the sea at slightly different times. Depending on the natural variability of the SST in space and time this could introduce spurious errors into the validation comparisons between AATSR and ISAR measurements of SST. Previous work on the consequences of this mismatch (Minnett, 1991) estimated for the north-east Atlantic that spatial separations of about 1 km and time intervals of about 2 h can introduce rms differences of 0.2 K between the mismatched samples. We can expect

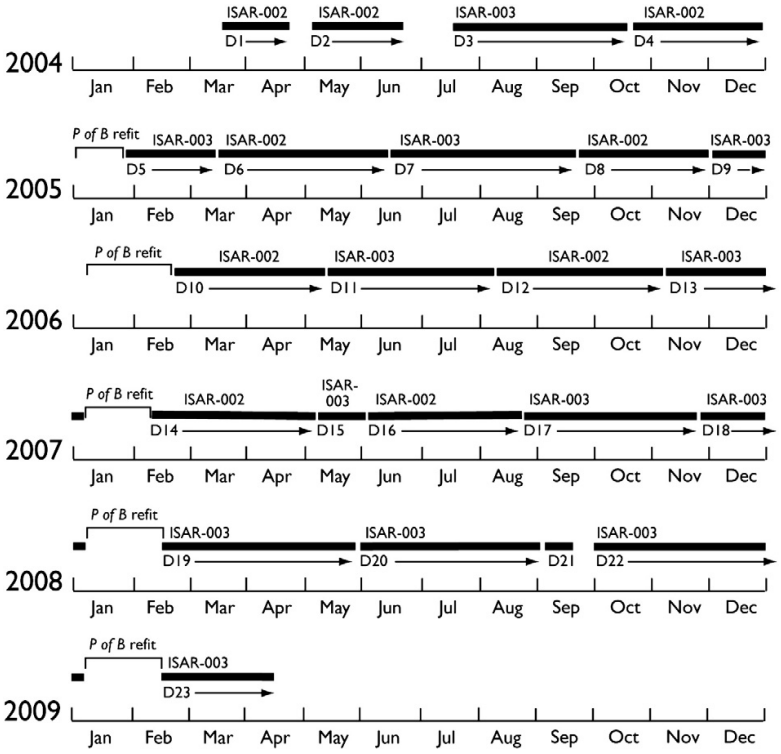


Fig. 2. Time-line showing when the ISAR system was operating on *Pride of Bilbao* during March 2004 to April 2009, which of the two instruments was deployed, and the deployment reference (D1 etc.) assigned to each deployment. No usable data were collected from D5, D9, D15 and D21.

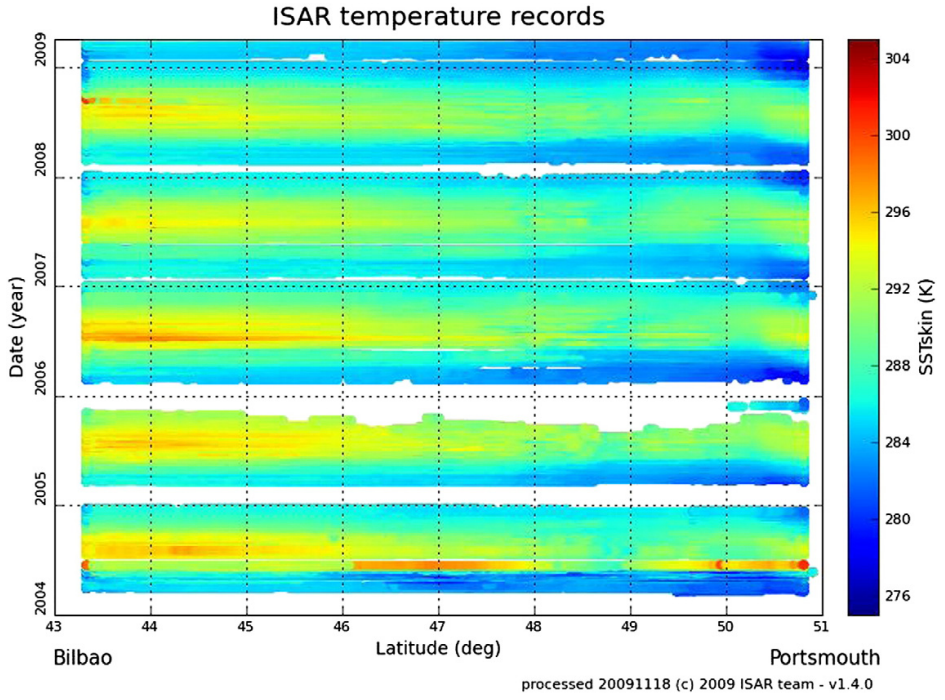


Fig. 3. Time-latitude plot of skin SST measured along the ship track by ISAR deployments between March 2004 and April 2009.

such differences to be unbiased. The magnitude of the error is also likely to vary considerably from place to place depending on the natural variability of the ocean and the local weather conditions.

In order to acquire the experience from which to identify the optimum matching criteria in the study area, we adopted an approach which specified five different grades of severity, defined in Table 1. Grade 1 corresponds to the closest coincidence considered feasible, within 1 km spatially and about half an hour in time, and effectively yields a match to the satellite image pixel containing the ship at the time of the overpass. Grade 2B limits the search radius to 1 km but allows a match to an earlier or later part of the ship track within a time window of 2 h either side. This is the matching criterion formally adopted by the AATSR Validation Protocol and thus it provides the core set of ISAR validation data used by the AATSR validation scientist. By Minnett's (1991) estimation it should lead to an rms mismatch of 0.2 K even if the ISAR and AATSR both recorded the true temperature. It should also be noted that the AATSR is supposed to represent the SST averaged over a pixel of area 1 km^2, whereas the ISAR samples the SST by integrating over 40 s of sea view in a measurement cycle that lasts 2 min. At typical ship speeds this results in one SST value averaged over a sample area of about 3×500 m, repeated at intervals of 1.5 km along the ship track.

Grade 2A allows a match to the cloud free pixel nearest to the ship track within a radius of 20 km but within about half an hour. Grade 3 allows matches with both the wider spatial search radius and the extended time window. This corresponds to the widest flexibility that we considered, at the start of the project, to be feasible in the English Channel/Bay of Biscay region. Grade 4 represents the coarsest of the criteria used by some agencies for open ocean validation of satellite SST data. It was included in order to discover whether such a wide match-up window can deliver useful validation results.

Fig. 4 illustrates some of the issues associated with identifying matched pairs of coincident samples of SST, one from a field of satellite image pixels (square, with side Δx) and one from a line of samples acquired from an underway shipboard device such as ISAR (for which Δt is the sampling interval, v is the ship speed and $v\Delta t$ is the spacing of the ISAR samples). Both cloud-free (a) and cloudy conditions (b) are considered. The limits of the coincidence criterion are also shown, using a search radius, $N\Delta x$, to constrain spatial coincidence and a time window, $\pm T$ centred on the overpass time, for temporal matching.

The preferred match-up in cloud-free conditions (shown in Fig. 4a) is between the ISAR sample obtained closest in time to the overpass (marked by the black dot) and the pixel (outlined and shaded grey) containing the ship at the time of the satellite overpass. However, for all the other ISAR samples (shown as open circles) obtained at times within the temporal sampling window of the overpass, a match is permitted to the pixel containing the geographical location where the ISAR sample was obtained. Thus a larger number of matched data pairs are possible for a line of successive samples along a ship track. For these to be genuinely independent no pixel and no ISAR record should be used in more than one validation match-up for a given overpass. In cloudy conditions, the search for matches can be extended across the whole search swath (see Fig. 4b) where matchmaking gives preference to minimising the spatial separation rather than the time difference, within the given coincidence window.

Table 1
AATSR–ISAR match-up grades and the relating match-up window size.

Grade	Match-up criteria
1	Coincidence of ISAR and AATSR sample within ±2000 s time window and 1 km search radius in space.
2A	Temporal match within ±2000 s and spatial match within ±20 km.
2B	Temporal match within ±2 h and spatial match within ±1 km.
3	Temporal match within ±2 h and spatial match within ±20 km.
4	Temporal match within ±6 h and spatial match within ±25 km

Software has been written to automate the validation matchmaking procedure which extracts matched pairs of AATSR and ISAR data following the successful completion of a deployment. A deployment is deemed to be successful once the post deployment validation against the CASOTS-2 black body has confirmed that the ISAR SST records are reliable to better than 0.1 K. The software sweeps through the ISAR data records twice. The first time it finds the AATSR overpasses containing cloud-free matches within the different coincidence criteria. The second time it refines the matches and eliminates duplicate entries of an ISAR record or AATSR pixel. It also identifies any matches which are geographically located within 5 km of the ports at either end of the ferry route. These are not included in the validation exercise, because they include ISAR records when the ship is berthed, which can be contaminated by the quayside or other vessels in the field of view, as well as SST records from congested and shallow port approaches where the SST is likely to be highly variable.

Since the objective of the exercise is to validate individual pixels of the AATSR SST data product, it is justifiable to obtain many match-ups from a single overpass as long as neither an AATSR pixel nor an ISAR SST record is used in more than one validation pair. Every ISAR SST record is acquired from an average of several detector counts and has its own unique set of black-body calibration samples, making it independent of the next ISAR record. Similarly, every AATSR pixel has undergone separate cloud detection tests and produces an SST using a unique set of brightness temperatures from the appropriate channels. It can therefore be argued that each validation pair, selected objectively in the way outlined above, provides an independent test of the validity of an AATSR-derived SST product when compared with a near-coincident ISAR measurement of SST.

It is important that a comprehensive validation exercise should aim to test the satellite data product not only over as large a number of pixels as possible, but also in a diverse range of conditions, especially if the atmospheric correction process is sensitive to those conditions. Since all the matched pairs between AATSR pixels from a single overpass and ISAR records along a single segment of ship track are likely to share similar atmospheric conditions as well as similar absolute SST and sea state, it is desirable that the entire dataset of matched pairs should come from a large number of separate overpasses, and should also span a wide range of SST values. The particular overpass is therefore recorded for each entry in the AATSR–ISAR match-up database, and the number of different overpasses represented in the match-up population is reported in the validation statistics. The range of SST values represented in the match-up data is also reported.

As far as possible, we have sought to produce a processing system which operates automatically without human intervention. Thus given a file or files containing the calibrated record of an ISAR deployment, and access to the AATSR data files covering the same period of time as the deployment, the programme must find the matched pairs for different grades of coincidence criteria, and then evaluate the mean and standard deviation of the differences for all matched pairs corresponding to a particular grade. There are two good reasons for using an automatic system. The first is that, once the system has been developed and validated, it significantly reduces the effort involved, leads to much more rapid matching of ISAR to AATSR data, and thus facilitates near-real time provisional monitoring of AATSR performance as long as the ISAR data can be transmitted to the project base at NOCS soon after acquisition. Secondly it avoids the need for subjective decisions to accept or reject match-ups, which might bias the matching and thus the validation process. The only filtering that has been applied to the datasets of AATSR–ISAR differences is to exclude from the bias evaluation any points lying outside three standard deviations of the set of observations being analysed.

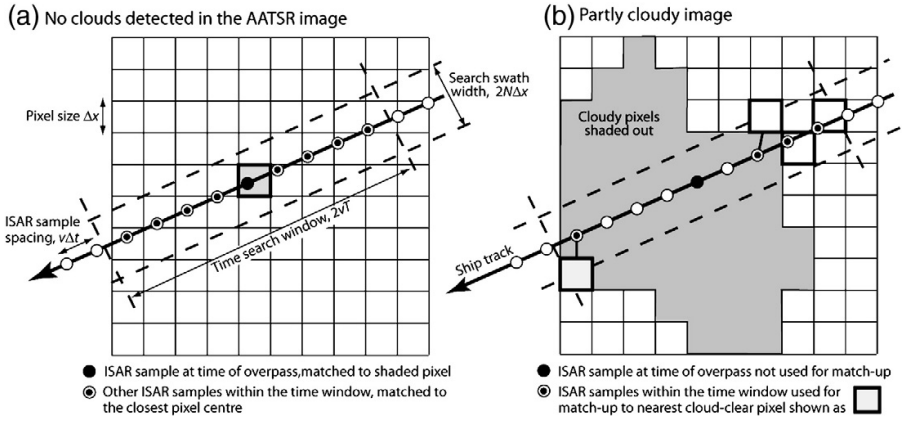


Fig. 4. Examples of match-up situations encountered in the construction of a match-up database using an in situ sensor moving at a speed *v*. (a) Cloud free conditions. (b) When some pixels are cloud contaminated.

4. ISAR–AATSR validation results

In this section, the differences between the AATSR SST retrievals (that is the consolidated data products derived from version 2 of the AATSR processor) and coincident ISAR data records are summarised across the whole five year dataset from March 2004 to April 2009. It is not appropriate to fully amalgamate all the data into a single dataset because changes were made to the AATSR algorithm coefficients, the new coefficients becoming operational from December 7th 2005. At the time of writing no revised SST products, reprocessed using the new coefficients, were available for AATSR data acquired before December 7th 2005. Instead we apply the validation in two stages, before and after the algorithm change. This also provides an opportunity to determine what effect, if any, the algorithm change made to the quality of SST retrievals. The discussion of the statistics focuses on the Grade 2B coincidence criteria, since these represent the match-up data used formally for the AATSR Validation Programme.

It is also necessary to consider separately the validation of the four different SST products from AATSR; Dual-2, Dual-3, Nadir-2 and Nadir-3. Dual-3 and Nadir-3 are available only at night when solar reflection is not a problem for the 3.7 μm waveband. Although Dual-2 and Nadir-2 are usable both day and night, we restrict the validation analysis to the daytime use of these products. In practise, since at night most users are expected to use the Dual-3 and Nadir-3, the Dual-2 and Nadir-2 products are not routinely generated unless the 3.7 μm waveband is unavailable.

4.1. Dual view algorithms

The primary SST product of the AATSR is that retrieved from the dual view. The comparison between the SSTs retrieved from the AATSR Dual-2 and Dual-3 algorithms and the coincident ISAR measurements is shown in Table 2 for the first phase up to 7th Dec 2005. Table 3 presents the match-up statistics for Dual-2 and Dual-3 data since 7th Dec 2005.

These confirm extremely good agreement between the in situ and the satellite temperatures. The magnitude of the bias is less than 0.1 K for the Grade 2B matches for both the Dual-2 and Dual-3 products during both phases, which means it is effectively indistinguishable from zero at the level of accuracy to which the ISAR itself is characterised. Comparing the biases before and after the algorithm change shows very little difference. For the three-waveband algorithm the slight positive bias shifts to a slight negative bias after the algorithm change, but for the two waveband algorithm there is no significant change to the slight negative bias.

The standard deviation about this very low mean bias is also small, although more variable with the match-up grade. For the Grade 2B match-ups that are used in the AATSR validation programme, the values are around 0.32 K for Dual-3 and 0.45 K for Dual-2.

The validation locations for the Grade 2B match-ups are well distributed along the line of the ship track, as shown in Fig. 5. Not all match-ups are shown. Each dot represents a unique overpass and is placed at the ISAR record location of the matched pair closest in time to the overpass. Because the ship's regular passenger timetable is phase-locked to the daily cycle in the same way as the sun-synchronous orbit of Envisat, there is a tendency to sample day and night overpasses in separate parts of the ship track, although there is some overlap. There is no reason to expect that this should influence the bias measurements in any way. However, since the region off Ushant where the SST is spatially more variable is normally traversed during the day, this may account for the higher variance from the Dual-2 match-ups for Grades 2A, 3 and 4 where the spatial coincidence tolerances are 20–25 km.

The scatter plots of the dual view validation pairs are shown in Figs. 6 and 7. These demonstrate graphically how good the matches are for both the Dual-2 and Dual-3 retrievals, and show the wide spread of temperatures over which they extend.

Table 2
Statistics for match-ups (AATSR dual-view SST-ISAR) from March 2004 to 7th December 2005, showing the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the range of sea temperatures spanned by the match-up database.

Grade of coincidence	Mean bias, AATSR–ISAR	Standard deviation	No. of matches	Overpass numbers	Min temp, °C	Max temp, °C
2 Waveband algorithm (Dual-2)						
1	−0.11	0.38	78	13	7.5	19.2
2A	−0.02	0.54	88	24	7.5	19.7
2B	−0.07	0.47	327	30	7.5	25.6
3	−0.06	0.55	385	42	7.5	25.6
4	−0.19	0.61	1154	56	7.5	25.6
3 Waveband algorithm (Dual-3)						
1	0.10	0.29	68	8	11.9	20.4
2A	−0.04	0.44	112	16	8.2	21.3
2B	0.05	0.32	232	16	7.8	20.9
3	0.00	0.37	425	27	7.3	21.6
4	−0.07	0.47	1428	58	7.3	22.5

Table 3
Statistics for match-ups (AATSR dual-view SST-ISAR) from 7th December 2005 to April 2009, showing the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the range of sea temperatures spanned by the match-up database.

Grade of coincidence	Mean bias, AATSR–ISAR	Standard deviation	No. of matches	Overpass numbers	Min temp, °C	Max temp, °C
2 Waveband algorithm (Dual-2)						
1	0.00	0.34	143	36	6.4	18.4
2A	−0.00	0.80	240	57	6.4	20.4
2B	−0.04	0.43	567	58	6.0	21.5
3	−0.03	0.64	843	68	5.1	22.0
4	0.04	0.84	2545	131	4.8	24.1
3 Waveband algorithm (Dual-3)						
1	−0.01	0.34	356	48	5.9	22.1
2A	−0.02	0.42	471	68	5.9	22.1
2B	−0.01	0.33	1227	62	5.9	22.6
3	−0.03	0.42	1607	87	5.9	22.6
4	−0.04	0.47	4945	140	4.1	24.9

4.2. Nadir-view algorithms

It is important to evaluate the performance of the SST products from the nadir-only algorithms, even though the results are less important in relation to the international role of AATSR to provide an absolute standard SST dataset for use as a climate reference. As in Section 4.1, before analysis of the match-up statistics the data are split into two time segments, before and after the algorithm change. These results are presented in Tables 4 and 5. Scatter plots for Grade 2B nadir-only match-ups are shown in Figs. 8 and 9.

Here the validation picture is different from the dual view retrievals. The nadir-3 algorithm performs well, comparable with dual-2 and dual-3. With reference to the Grade 2B and Grade 1 match-ups the nadir-3 retrievals seem to have improved from a bias of around 0.2 K before the algorithm change to a bias smaller than 0.05 K afterwards. In both cases the standard deviation is less than 0.25 K. Like the dual view retrievals this is a remarkable result that points to the fine performance of AATSR.

On the other hand, the nadir-2 algorithm shows a bias that is 0.5 K or more. Moreover, according to the Grade 2B matches, it has grown to 0.62 K after the algorithm change, although this difference may not be significant when viewed against the variability in the bias between different match-up grades. The high bias is a result of a problem with the nadir-2 algorithm that is well known to the AATSR validation team, and these data provide further confirmation of the problem, as well as a dataset against which to test any future algorithm changes to remedy it.

4.3. Effect on the validation of filtering out lower quality AATSR data

The GHRSST L2P version of the AATSR data products that are used operationally for global SST analyses and assimilated into ocean forecasting models (Donlon et al., 2007; Robinson et al., in press), not only contains SST fields identical to the standard dual-view products but also includes additional fields such as error statistics and a Confidence Value (CV). The objective measure used for assessing the errors on a pixel-by-pixel basis is the magnitude of the difference between SST estimated by the dual view algorithm and by the nadir-only view, referred to as the (D-N) difference. This approach, first proposed by Noyes et al. (2006) from the results of their analysis, was implemented and evaluated by Corlett and Poulter (2008) using an extended match-up database of AATSR SSTs compared with buoy SSTs. Pixels for which the value of (D-N) is found to be outside specified thresholds are assigned a lower CV of 3 (Marginal quality) or 4 (Good), instead of the highest value of 5 (Excellent). Table 6 defines these thresholds and how the CV is assigned. Operational users seeking the highest quality of data with the least uncertainty are recommended to select only those data for which CV = 5.

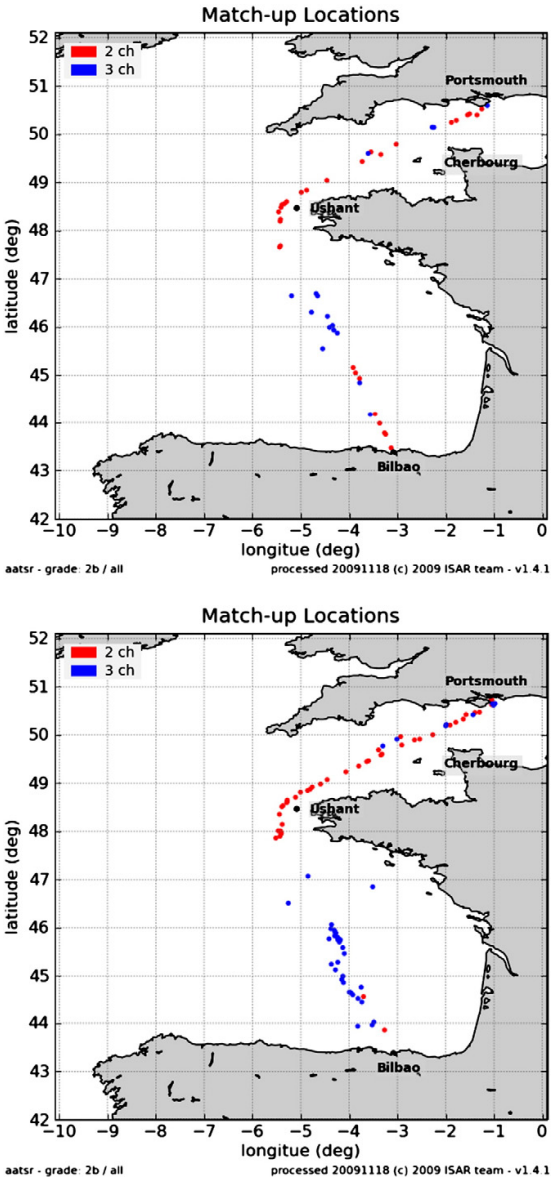


Fig. 5. Location of the match-ups at Grade 2b coincidence for AATSR dual view SST retrievals against ISAR observations from four years of deployments on Pride of Bilbao, March 2004 to December 2005 in the top panel and December 2005 to April 2009 in the bottom panel. The colours distinguish 2-waveband data (day: red) from 3-waveband data (night: blue).

So far the comparisons between AATSR and ISAR have included all the available AATSR data other than those assigned CV = 2 (bad quality). A revised match-up dataset was produced for all observations acquired since the AATSR algorithm change on 7 Dec 2005. This was achieved by evaluating (D-N) for every AATSR record in the match-up database, assigning a CV according to the thresholds shown

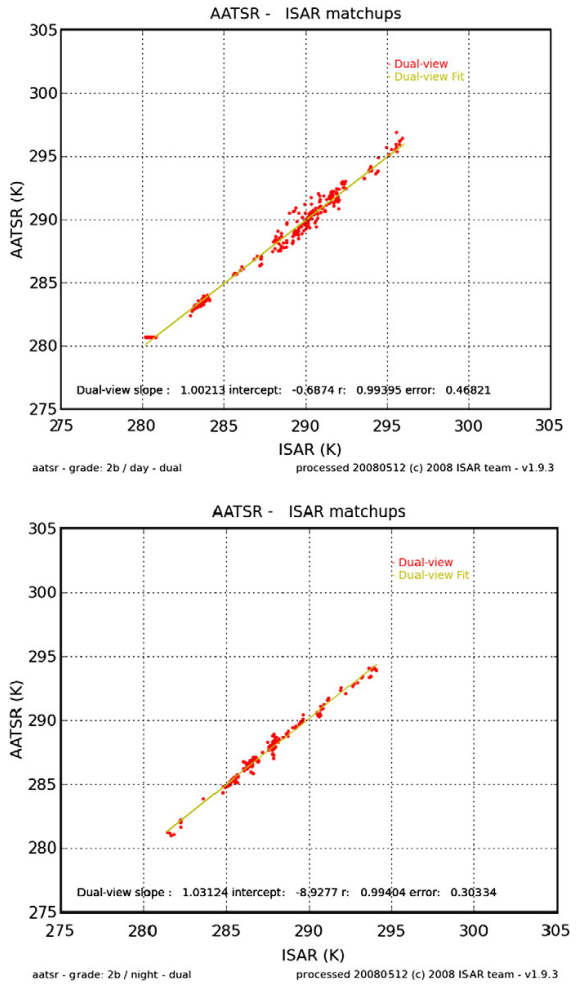


Fig. 6. Scatter plot of the AATSR-SST dual-view retrievals and the Grade 2b coincident ISAR observations, for the period March 2004 to December 7th 2005. Top panel shows the 2-waveband retrievals (day) and the bottom panel shows the 3-waveband retrievals (night).

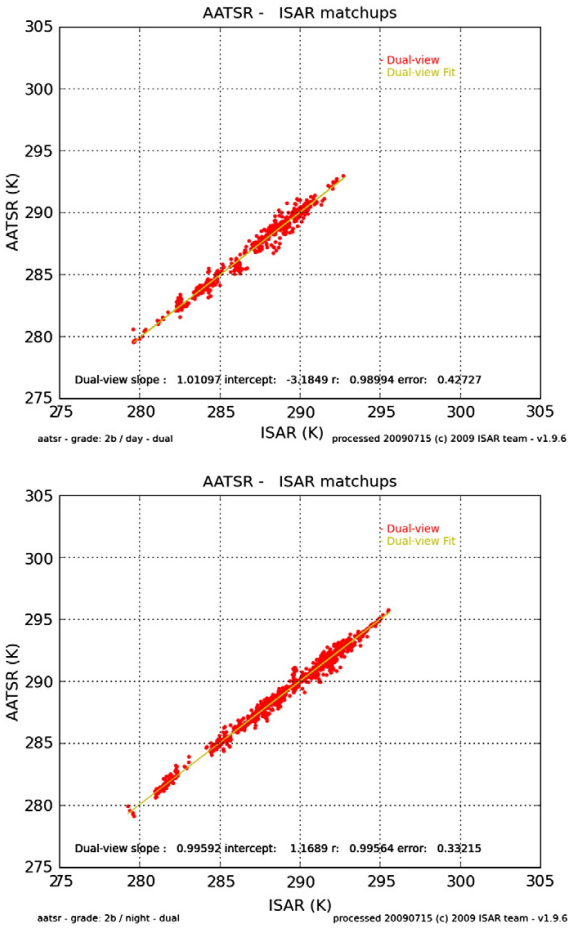


Fig. 7. Scatter plot of the AATSR-SST dual-view retrievals and the Grade 2b coincident ISAR observations, for the period December 7th 2005 to April 2009. Top panel shows the 2-waveband retrievals (day) and the bottom panel shows the 3-waveband retrievals (night).

in Table 6 and then removing any match-ups for which the CV of the AATSR record is less than 5. Table 7 shows the AATSR–ISAR match-up statistics with all lower quality AATSR data filtered out.

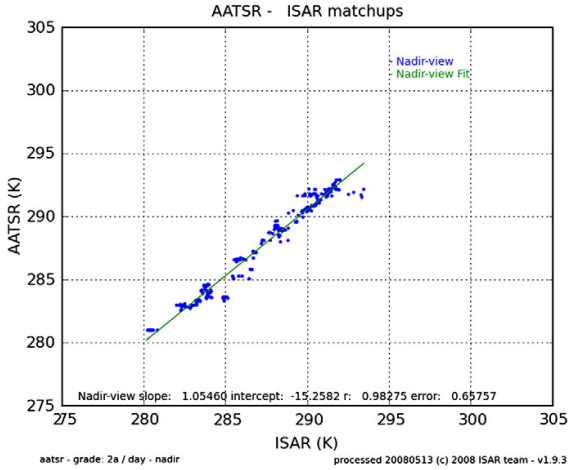
A comparison between Tables 7 and 3 reveals the impact of filtering by AATSR Confidence Value. Note first that only a small proportion of the matched data have been eliminated because the majority of the cloud-free AATSR pixels in this region are of excellent quality. The right hand column in Table 7 shows that the percentage of pixels removed is around 10% although for the Grade 2B matches it is only 3% for Dual-2 data and 7% for Dual-3 data. The filtering causes little change to the mean differences between AATSR and ISAR, which remain smaller in magnitude than 0.05 K. On the other hand, there is a consistent reduction of the standard deviation, in the case of the Grade 2B matches it reduces from 0.43 K to 0.39 K for Dual-2, and for Dual-3 it reduces from 0.33 to 0.26 K. There are similar, sometimes larger, reductions of standard deviation for the other grades of match-up, effectively demonstrating that eliminating pixels for which CV is less than 5 reduces the errors in the resultant SST datasets.

Table 4
Statistics for match-ups (AATSR nadir-only SST-ISAR) from March 2004 to 7th December 2005, showing the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the range of sea temperatures spanned by the match-up database.

Grade of coincidence	Mean bias, AATSR–ISAR	Standard deviation	No. of matches	Overpass numbers	Min temp, °C	Max temp, °C
2 Channel algorithm (Nadir-2)						
1	0.54	0.43	147	13	7.5	19.2
2A	0.43	0.68	287	24	7.5	19.7
2B	0.49	0.44	615	30	7.5	25.6
3	0.42	0.62	1165	42	7.5	25.6
4	0.37	0.62	3272	56	7.5	25.6
3 Channel algorithm (Nadir-3)						
1	0.23	0.21	100	8	11.9	20.4
2A	0.06	0.36	230	16	8.2	21.3
2B	0.20	0.24	332	16	7.8	20.9
3	0.03	0.42	805	27	7.3	21.6
4	−0.02	0.49	2903	58	7.3	22.5

Table 5
Statistics for match-ups (AATSR nadir-only SST-ISAR) from 7th December 2005 to April 2009, showing the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the range of sea temperatures spanned by the match-up database.

Grade of coincidence	Mean bias, AATSR–ISAR	Standard deviation	No. of matches	Overpass numbers	Min temp, °C	Max temp, °C
2 Channel algorithm (Nadir-2)						
1	0.55	0.38	298	36	6.4	18.4
2A	0.56	0.66	655	57	6.4	20.4
2B	0.62	0.44	1138	58	6.0	21.5
3	0.58	0.64	2211	68	5.1	22.0
4	0.68	0.85	7358	131	4.8	24.1
3 Channel algorithm (Nadir-3)7						
1	0.05	0.29	603	48	5.9	22.1
2A	−0.01	0.36	895	68	5.9	22.1
2B	0.05	0.25	2077	62	5.9	22.6
3	−0.02	0.38	3246	87	5.9	22.6
4	−0.05	0.45	10,555	140	4.1	24.9



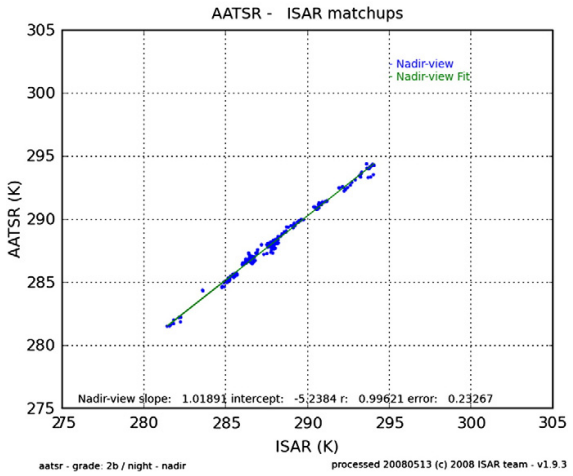
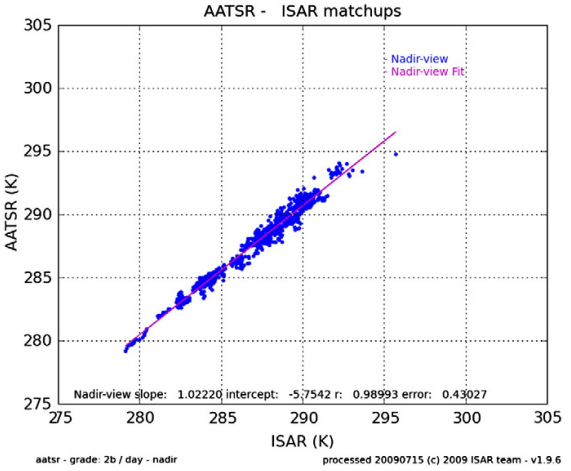


Fig. 8. Scatter plot of the AATSR-SST nadir-only retrievals and the Grade 2b coincident ISAR observations, for the period March 2004 to December 7th 2005. Top panel shows the 2-waveband retrievals (day) and the bottom panel shows the 3-waveband retrievals (night).



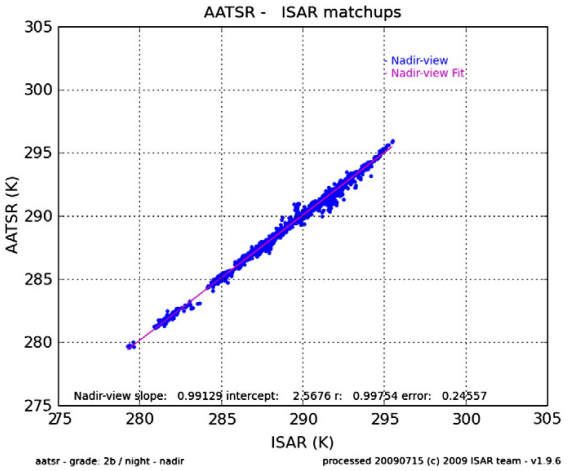


Fig. 9. Scatter plot of the AATSR-SST nadir-only retrievals and the Grade 2b coincident ISAR observations, for the period December 7th 2005 to April 2009. Top panel shows the 2-waveband retrievals (day) and the bottom panel shows the 3-waveband retrievals (night).

4.4. Histograms of the match-up differences

Fig. 10 shows the histograms for the dual-view (red) and nadir-only view (blue) differences between AATSR and ISAR before the algorithm change on December 7th, separately for 2-channel and 3-channel retrievals for three grades of match-up coincidence between the satellite and in situ radiometer measurements. Fig. 11 presents the

Table 6
(D-N) temperature difference thresholds for assignment of AATSR dual view confidence values (CV) of 5 (excellent quality), 4 (good quality) and 3 (marginal quality), as determined by Corlett and Poulter (2008).

	2 Channel algorithm	3 Channel algorithm
CV for (D-N) above upper threshold	3	4
Upper threshold temperature difference, K	0.04 K	0.51 K
CV for D-N between upper and lower threshold	5	5
Lower threshold temperature difference, K	−1.53 K	−0.51 K
CV for (D-N) below lower threshold	3	4

Table 7
Statistics for match-ups from 7th December 2005 to April 2009 (AATSR dual-view SST-ISAR), filtered to remove AATSR data with lower confidence values. The table shows the bias and standard deviation, the number of matches, the number of overpasses from which they came, and the percentage reduction in the match-up population compared with the unfiltered data as in Table 3.

Grade of coincidence	Mean bias, AATSR–ISAR	Standard deviation	No. of matches	Overpass numbers	Population reduction by removing data with CV<5
2 Waveband algorithm (Dual-2)					
1	−0.03	0.30	136	36	5%
2A	−0.02	0.60	209	57	13%
2B	−0.04	0.39	549	58	3%
3	−0.05	0.54	784	68	7%
4	0.02	0.76	2372	131	7%
3 Waveband algorithm (Dual-3)					
1	0.00	0.23	326	48	8%
2A	0.00	0.32	411	68	13%
2B	−0.01	0.26	1137	62	7%
3	−0.03	0.32	1423	87	11%
4	−0.05	0.38	4299	140	13%

same set of histograms for all match-up data after 7th Dec 2005. The comparison between these two figures provides a more complete view of any changes to the character of the SST products in response to the algorithm coefficient changes.

The effect of filtering the dual-2 and dual-3 data products using the Confidence Value is illustrated in Fig. 12, where the upper two rows show the histograms of the unfiltered AATSR–ISAR temperature differences for four different grades of match-up, and the lower two rows repeat these for the filtered case. Comparison between the different columns of this figure offers an insight into how the different match-up criteria affect the validation dataset.

4.5. Year-by-year match-ups

In order to test the stability of the AATSR dual view SST product performance, the match-up statistics were evaluated for individual years following the algorithm coefficient change. The data span the three complete years 2006, 2007 and 2008. The results are shown in Table 8, for Grade 2B matches only. Given the smaller population sizes within a single year, some variation of the bias from year to year is expected, but it turns out to be very small, spanning 0.09 K for Dual-2 products and 0.05 K for Dual-3, both within the absolute calibration limit of the ISAR.

5. Discussion

5.1. The validation of AATSR SST products using in situ radiometry

The primary objective of the programme of work presented in this paper was to deliver an independent and objective assessment of the accuracy of the SST data retrieved from the AATSR, with particular attention to the dual-view products which are increasingly being used operationally in combination with other satellite SST data. The SST data from AATSR are derived to represent the skin temperature of the ocean. By making in situ measurements of the very same skin temperature using the ISAR shipborne radiometer, it has been possible to create a match-up dataset from which the accuracy of AATSR can be directly evaluated relative to the independent ISAR

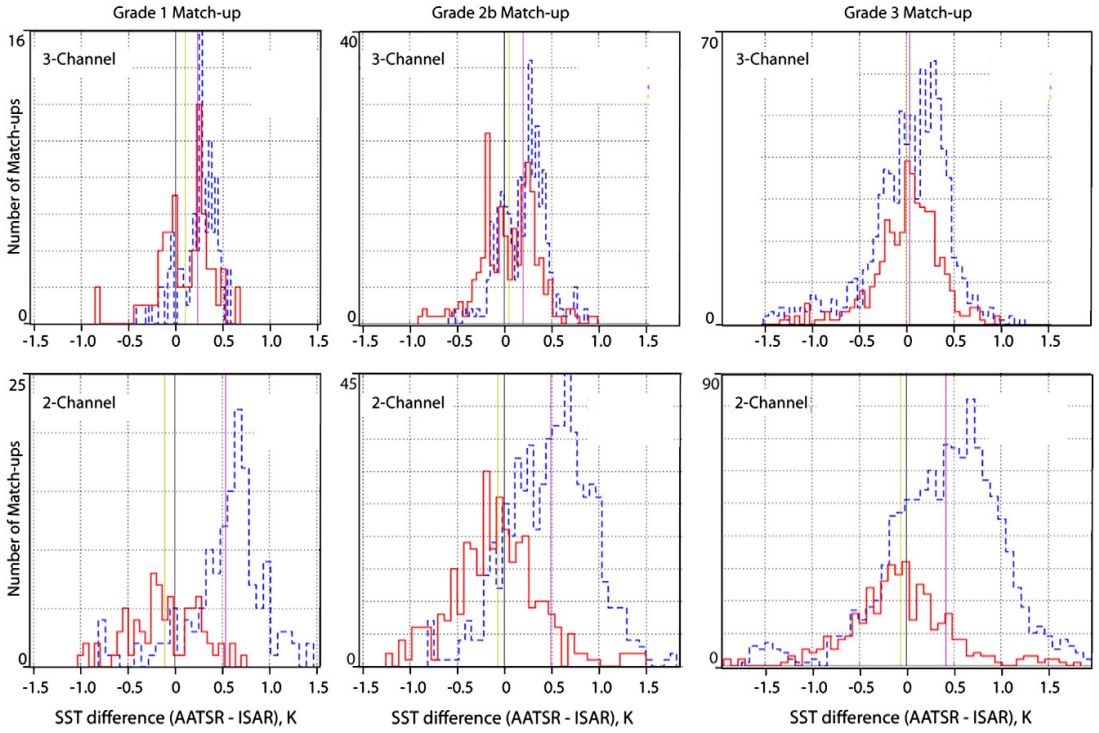


Fig. 10. Histograms of the distribution of the difference between the AATSR and ISAR temperature measurements during the period March 2004 to Dec 7th 2005, for the dual view (red, continuous line) and nadir-only (blue, dashed line) retrievals. The bin size is 0.07 K. The upper row refers to 3-channel AATSR data, the lower row to 2-channel. The first column contains Grade 1 matches, the second is for Grade 2b matches and the third for Grade 3.

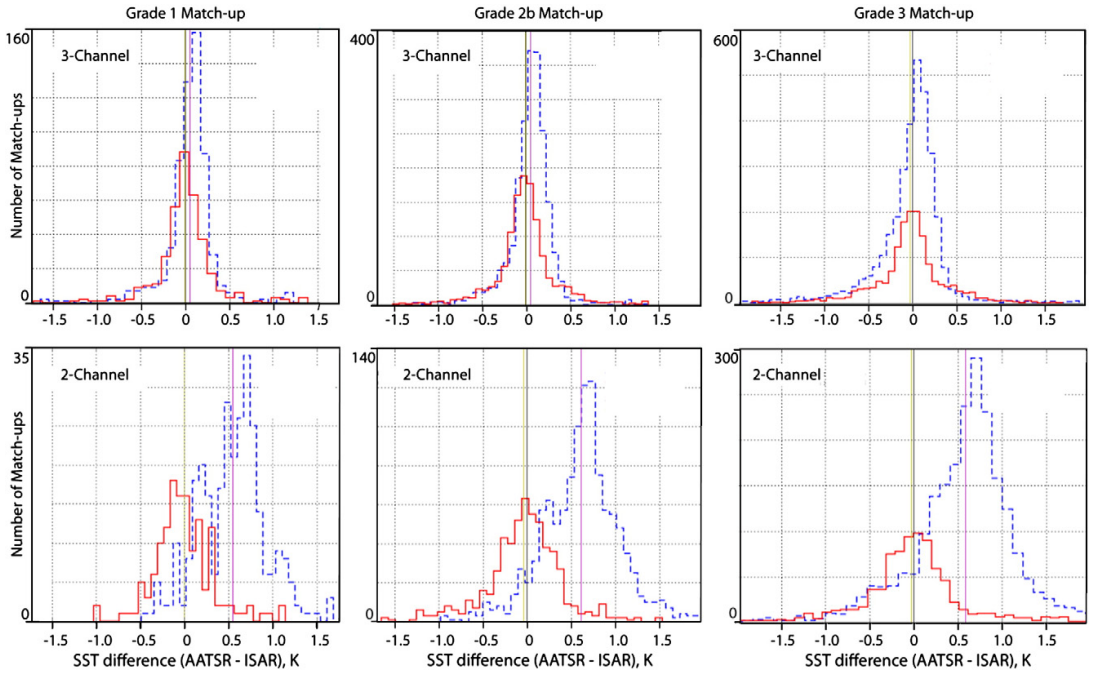


Fig. 11. Histograms of the distribution of the difference between the AATSR and ISAR temperature measurements during the period Dec 7th 2005 to April 2009, for the dual view (red, continuous line) and nadir-only (blue, dashed line) retrievals. The bin width varies with the total population, between 0.07 and 0.1 K. The upper row refers to 3-channel AATSR data, the lower row to 2-channel. The first column contains Grade 1 matches, the second is for Grade 2b matches and the third for Grade 3.

measurements. The results presented in Table 7 show that the difference between the AATSR dual-3 product (CV = 5) and in situ measurements made within 1 km and 2 h of the overpass is -0.01 ± 0.26 K (from 1137 independent samples), and for the dual-2 product (CV = 5) the difference is -0.04 ± 0.39 K (from 549 samples). This represents primary evidence of the excellent performance, during the period Dec 2005 to Apr 2009, of the AATSR measuring system for SST, including the sensor itself and the SST retrieval algorithms which cope with the variable effects of atmospheric interference on the infrared radiation reaching the satellite. These validation results are consistent with those obtained in the Caribbean (Noyes et al., 2006) using a different shipborne radiometer (0.02 K ± 0.25 K from 74 samples for Dual-3, and −0.05 K ± 0.26 K from 25 samples for Dual-2). Moreover, the year-by-year statistics in Table 8 confirm the underlying stability of the AATSR dual-view SST products, although it remains desirable to apply this test across the full span of coincident AATSR data (when it has all been reprocessed with the same algorithm coefficients) if it is to be confirmed that there are no significant trends in the bias of AATSR relative to ISAR.

We can identify three factors other than the behaviour of the AATSR itself which influence the match-up statistics and which should be taken into account when objectively assessing the overall validity of AATSR data. The first is the performance of the ISAR sensor. The operational regime for the deployment of ISAR is designed to ensure that it delivers sea surface skin temperature measurements with an absolute accuracy within ± 0.1 K, confirmed by regular characterisations of its performance against independent references traceable to international standards.

Secondly, as evidenced by the tables of match-up statistics, the comparisons between satellite and in situ data vary with the criteria used to define "coincidence" in space and time between the two samples of temperature, as discussed in Section 5.3.

The third factor is that the validation results presented here correspond to a limited geographical region covering the Bay of Biscay and the English Channel. Because this is a region which has a fairly stable atmosphere and almost no latitudinal retrieval bias (Merchant et al., 2008) it would be unwise to assume that the AATSR performance is necessarily of equal reliability over other parts of the ocean where the atmospheric effects may be more complex. This highlights the need to extend the geographic coverage of shipborne infrared radiometers.

Unlike most validation studies for satellite-derived SST, we can be sure that the results presented here contain no uncertainty associated with near surface variability of the ocean's thermal structure caused by the thermal skin or by diurnal warming effects. Most other published AATSR validation results have had to use in-situ measurements made at some depth below the skin. Various strategies have been adopted to minimise the uncertainty this may introduce, for example by discarding day time data to avoid falsification of results by diurnal warming events during the day time, or by applying skin-to-bulk conversion algorithms to the AATSR SST and evaluating the comparison to buoy data as "bulk" SST (O'Carroll et al., 2008). Such skin to bulk conversions can introduce further error into the validation process which may degrade the apparent performance of the AATSR SST data.

It is inevitable that sub-surface contact SST measurements are used for most comparisons with satellite SST because at present there are far more buoy observations available from around the world, essential for a global product validation. For example Corlett et al. (2006) showed for a global AATSR analysis with buoy match-ups that the results for Dual-3 are 0.04 ± 0.28 K and for Dual-2 they are 0.02 ± 0.39 K. However, although these appear similar to the results of the AATSR–ISAR comparisons presented here, such comparisons are inappropriate because one is global and the other is regional. Corlett

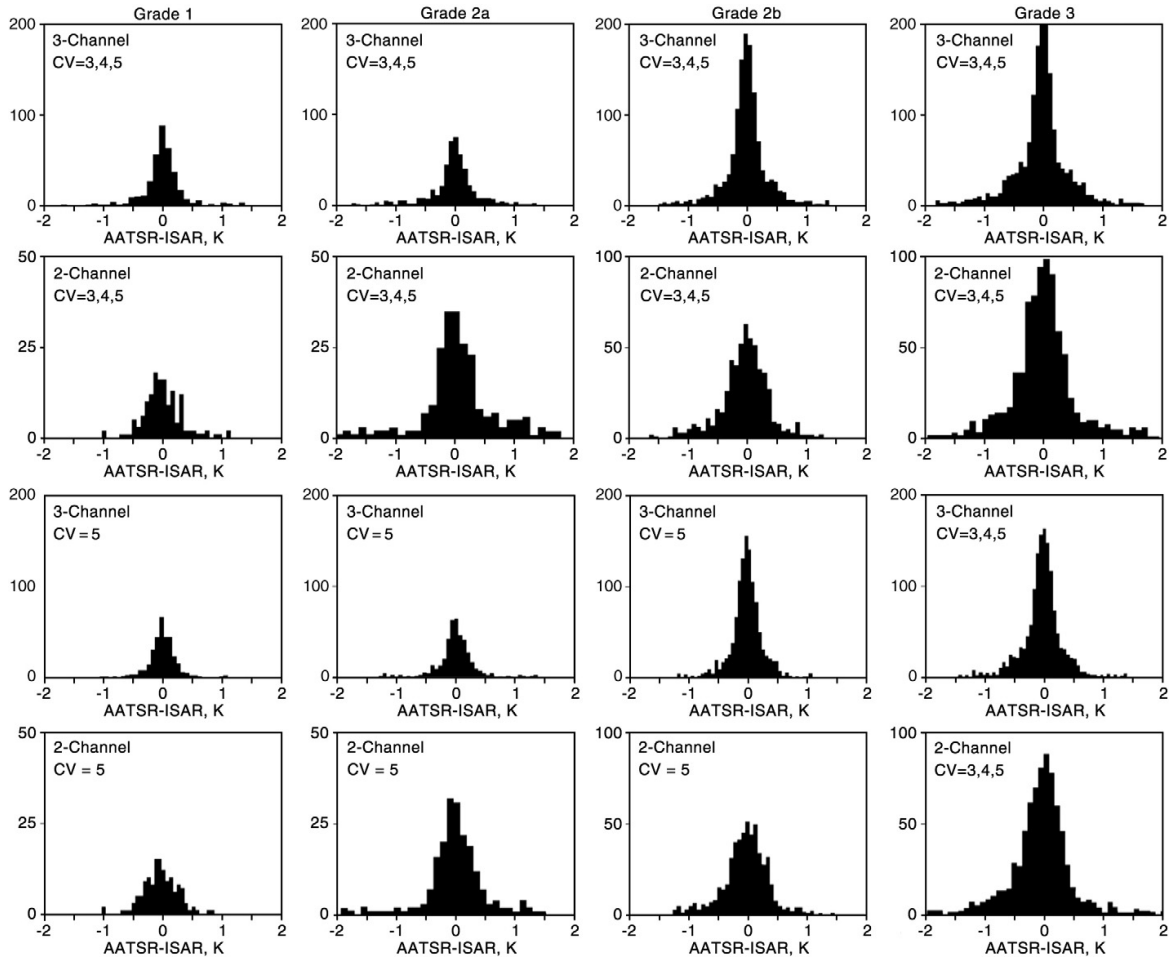


Fig. 12. Histograms of the distribution of the difference between the AATSR dual-view SST and ISAR temperature measurements during the period Dec 7th 2005 to April 2009. The upper two rows include AATSR data of at least acceptable quality, whereas the lower two rows contain only AATSR data of highest quality (Confidence value = 5). Lines 1 and 3 contain dual-3 data, lines 2 and 4 are for dual-2 data. The first column contains Grade 1 matches, the second is for Grade 2a, the third for Grade 2b and the fourth for Grade 3.

et al. (2006) concluded that, because direct comparison between AATSR and buoy data should include the cool skin average bias of −0.17 K, they were in practise detecting a warm bias of about 0.2 K in the global match-up that called for further regional analyses. The results of this paper show with some confidence that, in the Biscay area, there is no appreciable warm bias of the AATSR dual-view SST products. It would be useful to be able to provide similar, unambiguous, shipborne radiometer data for other regions of the ocean in order to explore whether there is a regionally variable bias in the AATSR data.

Corlett et al. (2006) also included a comparison between AATSR and skin temperatures from four different radiometer campaigns in different global regions (M-AERI, SISTER, DAR011 and ISAR, see (Barton et al., 2004) for information on the other radiometers). The combined results showed a difference of 0.06 ± 0.2 K for Dual-3 and 0.11 ± 0.33 K for Dual-2, which are slightly warmer than the results presented here, although the sample set used by Corlett et al. was quite small with only 30 independent match-ups.

Table 8
Statistics by year for 2006, 2007 and 2008 of Grade 2B match-ups (AATSR dual-view SST-ISAR), filtered to remove AATSR data with lower confidence values.

Year	Mean bias, K	Standard deviation, K	No. of matches	Overpass numbers
2 Channel				
2006	0.07	0.49	92	17
2007	0.00	0.33	247	23
2008	−0.02	0.49	110	12
3 Channel				
2006	−0.02	0.26	220	14
2007	−0.02	0.21	463	22
2008	0.03	0.39	308	17

5.2. Monitoring the outcome of data product changes

From the outset of the reported study its limited geographical coverage was acknowledged, but this was traded against the advantage of obtaining a long term record of in situ observations from the same platform and sensor type, over the same regional sea. The five year dataset from ISAR has revealed no obvious long term trend in the comparisons with coincident AATSR data. It therefore provides a useful set of data against which to evaluate the impact of

changes made to the coefficients used in the AATSR retrieval algorithms, and also to evaluate the consequences of applying filters based on the AATSR Confidence Value.

Inspection of the histograms in Figs. 10 and 11 adds further insight to the information contained in Tables 2–5, about the response of the AATSR retrieval characteristics following the change of algorithm coefficients on 7th Dec 2005. Tables 2 and 3 confirm that both the AATSR dual view algorithms produced very good results before the coefficient change on December 7, 2005, and that a small improvement, (reduction of biases) was gained by making that change. Although the standard deviation did not change for Dual-3 and Dual-2, the histograms (solid red line) in Figs. 10 and 11 show a much more regular Gaussian-like shape following the coefficient change than before. This must be due partly to the increased number of match-up samples available after 7th Dec 2005, but it also seems to be the case that more match-up samples are found close to zero difference, producing a better defined central peak to the histogram. The standard deviation does not change much because there are a similar number of outliers (beyond ± 0.5 K) in each case.

The improved match-ups show that the Nadir-3 (night) SST products benefit greatly from the coefficient change, making their statistics (see Table 5) as good as those of the Dual-3 match-ups. The evidence of this is portrayed graphically by the blue dashed lines in the top row of Fig. 11 where the distribution of AATSR–ISAR differences is seen to be very sharply peaked around zero, and aligned with the peak of the Dual-3 data, for each of the different Grades of match-up. This contrasts with the top row of Fig. 10 where the data from before the algorithm coefficient change appear to be much less tightly clustered, giving rise to biases which vary by up to 0.2 K between different match-up Grades, although there is little difference between the standard deviation of the matches from before and after the algorithm coefficient change.

The Nadir-2 results (daytime data) show a large warm bias which does not change substantially with the coefficient change. It appears in Fig. 11 that the offset peak of the Nadir-2 histogram is better defined after 7th Dec 2005 but its width is still very much greater than for the other three SST products, giving rise to the large standard deviation associated with the match-ups for this SST product.

Fig. 12 isolates the dual view products from the others to compare the histograms when all Confidence values (3 to 5) are included (upper two rows) and those when only excellent quality data are retained (lower two rows). The histograms in the first case already have well defined peaks. This does not change following the removal of the lower quality data, whilst the distributions become narrower and more Gaussian. In addition, where coincidence within 1 km is required the number of outliers (outside about ± 0.8 K) is greatly reduced for the Grades 1 and 2B matches (Columns 1 and 3). This upholds the validity of the Confidence values and supports the practise of filtering out lower quality data for applications where individual pixel values need to be as accurate as possible.

5.3. Under-way sampling from ships and the match-up window size

Dedicated campaigns that are specially mobilised for calibrating or validating ocean data products from satellites are normally limited in duration and typically suffer from the difficulty of acquiring sufficient in situ data points matching satellite data to make the campaign worthwhile. This is especially problematic for infrared SST measurements where cloud cover hides a large proportion of overpasses, and it is seriously exacerbated in the case of the AATSR by the narrow (500 km) swath of that sensor resulting in a revisit interval (day or night) of longer than one day at most latitudes. In this context the use of unattended sensors on a ship of opportunity with nearly continuous operation for several years greatly increases the number of matches between in situ and satellite observations, provided the vessel is at sea at the times when the satellite overpasses. In the case of the *Pride of Bilbao*'s normal schedule it is in port for only 3 out of every 10 overpasses. However, it is sometimes possible to find matches even for those overpasses for which the ship is in port when the satellite is overhead, if a match-up time window of ± 2 h is allowed. This can occur if the ship is at sea sometime within the ± 2 h window. Moreover, for the 7 out of 10 overpasses when the vessel is at sea, the ship travels several tens of km during such a window and this increases the number of independent match-ups available from any overpass that has cloud-free pixels in the vicinity of the ship.

The number of validation match-ups acquired over the five year period of this study provides compelling evidence of the effectiveness of deploying an autonomous radiometer for this purpose on a ship of opportunity. Taking Grade 2B as the match-up criterion required by the AATSR Quality Assessment Panel, a total population of 2353 matches between ISAR and AATSR have been acquired from 166 different overpasses. This allows a much more thorough testing of the AATSR performance than was possible in the early life of AATSR when only a few isolated skin SST measurements were available. Even this number is not very large for a five year period, which serves to highlight the difficulty of acquiring validation data for a sensor with such a narrow swath.

Varying the match-up window has a large effect on the number of independent matches, as shown in Tables 2–5. However, widening the spatial window from 1 km to 20 km (e.g., from Grade 1 to Grade 2A) increases the number of matches by a lot less than extending the time window from ± 2000 s to ± 2 h. (Grade 1 to Grade 2B). Whilst the preferred use of Grade 2B stems from previous published work (Minnett, 1991) we can use the match-up dataset acquired here to explore how the size of the window degrades the match-up statistics. Degradation arises by allowing cases where the ISAR and the AATSR were pointing at different SSTs, whilst making their measurements, to be compared as "coincident". Here the histograms in Fig. 12 are useful, alongside the tables of match-up statistics. Focussing on the highest quality AATSR data (Table 7 and the lower two rows of Fig. 12) it is evident that Grade 1 delivers the lowest standard deviation and the cleanest histograms with a minimum of outliers. However, the population for Grade 1 is smallest and the resulting histograms tend to be less smooth than for the other grades. In order to achieve the most reliable estimate of the mean bias it is desirable to use a larger population. Relative to Grade 1, the Grade 2A histograms tend to have a wider spread of outliers, and the main histogram shape is broadened, implying that widening the spatial window has introduced false discrepancies into the match-up data, with only a small increase in population. On the other hand the Grade 2B distributions have fewer outliers than Grade 2A, but with a much larger population gain, and a well defined histogram peak. This implies that widening the time window is a more effective way of enhancing the population with less degradation of the match-ups than widening the spatial window. The Grade 2B criterion seems to be a satisfactory compromise. Grade 3, with both space and time windows widened, contains more outliers and leads to a wider histogram although it seems to preserve a symmetrical shape from a large population which ought to provide a reliable bias estimate. The Grade 4 histogram is not shown but it is wider still, with a larger standard deviation. Whilst it might be acceptable in the open ocean, in waters which have strong tidal currents a 6 h window is not considered acceptable.

The increasing standard deviation with widening match-up window is a warning that even though we are comparing identical types of SST (that is the skin temperature) there are still spurious mismatches introduced because of the natural variability in the ocean coupled with the difficulty of ensuring that the in situ and the satellite sensors are precisely observing the same patch of sea. This implies that the reported standard deviations are not a fair representation of the AATSR's performance. It can be conjectured that the standard deviation would reduce further if it were possible to eliminate matched pairs that are not actually comparing like with like. This highlights the challenge to develop matching techniques that take account of the factors that

introduce spurious mismatch errors, which have nothing to do with the performance of either the AATSR or the ISAR. Only when that is achieved will it be possible to use the standard deviation of comparisons against ISAR data to discriminate between the quality performance of AATSR and that of SST products from other satellites.

6. Conclusion

The ISAR validation campaign has provided an operational SST skin measurement programme for more than five years and has demonstrated its importance in the quality control of AATSR during that time. Validation of AATSR against the ISAR instrument has confirmed the excellent performance of the SST data retrieved using the Dual-3 and Dual-2 algorithms. The measured differences between the satellite and ship radiometer retrievals of these products are -0.01 ± 0.26 K and -0.04 ± 0.39 K respectively. Since the ISAR has an absolute accuracy within ± 0.1 K this inspires confidence in the use of AATSR dual-view SST data, firstly as a suitable bias-correcting reference against which other satellite SST products can be adjusted when combined for operational applications, and secondly as the basis for producing a climate quality SST record. However the accuracy and stability of the AATSR reported here is strictly applicable only to the limited geographical coverage and temporal extent of this study. Moreover it is essential to sustain ongoing validation programmes of this nature to ensure that the satellite instrument performance is maintained and the accuracy of its SST products are validated throughout the complete data record.

The use of well calibrated radiometry from a ship of opportunity has proved to be effective in reliably measuring skin SST. The low biases and standard deviation in the match-up dataset have demonstrated the advantage of removing the uncertainty that arises in conventional validation programmes when satellite skin SST measurements are compared against in situ measurements of subsurface temperature. The capacity of an autonomous radiometer to sustain extended high quality deployments without the need for frequent operator intervention has been proven. The ability to acquire a few hundred match-up points per year with a single ship instrument provides a dataset suitable for quality analysis and algorithm refinement purposes. There remains a challenge to improve the matching procedure between in situ and satellite records to avoid the introduction of spurious discrepancies.

Acknowledgements

This work was funded by the UK Department of Environment, Food and Rural Affairs (Defra) and Department for Energy and Climate Change (DECC) through Contract No. 2004-03-002/CPEG 10, managed by Space ConneXions Ltd. We acknowledge the generosity of P & O plc in allowing the ISAR instrument to be deployed on the *Pride of Bilbao*, and wish to record our thanks to the Masters and crew of that vessel since 2004 for their ongoing assistance. The European Space Agency provided ready access to all the AATSR data.

We are grateful to Hugh Kelliher of Space ConneXions Ltd, Gary Corlett, David Llewellyn-Jones and John Remedios of University of Leicester, and Tim Nightingale of Rutherford Appleton Laboratory, partners in the AATSR Data Exploitation Contract, for ongoing constructive criticism of this work. We also acknowledge the outstanding technical support provided by Gary Fisher, the late Ray Collins and colleagues in the workshop of the School of Ocean and Earth Science in the National Oceanography Centre, Southampton.

Appendix A. Modification to the ISAR internal calibration procedure

The purpose of this appendix is to define a modification made to the previously published description of the internal calibration procedure used to retrieve estimates of SST using the ISAR. The ISAR operational cycle involves a 30 s integration of the detector measuring radiation from the internal ambient temperature black body, then a 40 s integration of the sea view, then a 10 s integration of the sky view and finally a 30 s integration of the heated internal black body radiation. The raw counts from the sea and sky views are converted into equivalent brightness temperatures using the views of the two blackbodies internal to the instrument during the same measurement cycle, along with the associated thermistor temperatures. The SST retrieval follows the procedure defined in Donlon et al. (2008) which was intended to automatically adjust for degradation of the optical transmission path through the instrument, particularly the scan drum mirror. However, early experience showed that the procedure needed refinement to allow for the additional scattering by the scan mirror as it slowly degrades during a three month deployment.

Although the original procedure automatically adjusted the calibration to allow for the reduced transmission by the degraded optical path, it did not allow for the slight increases in the proportion of stray (ambient) radiation reaching the detector because of the additional diffuse scattering by the scan mirror. This affects the estimate of the absolute radiance, L_{BB}, reaching the detector when the scan mirror points towards each of the black bodies, which is needed within the main SST retrieval procedure. This is defined as:

$$L_{BB} = \varepsilon B_B(T_{bb}) + (1-\varepsilon)B_B(T_{amb}) \tag{A1}$$

where ε is the effective emissivity of the blackbody (viewed through the scan-drum aperture) and B_B is the Planck function integrated over the detector bandwidth, evaluated for the blackbody temperature (T_{bb}) or the ambient temperature (T_{amb}) internally within the ISAR, both of which are measured to a high accuracy by thermistors.

The effective emissivity is a measure of the non-blackness of the blackbody integrated across the field of view of the radiometer through the scan-drum aperture. This may encounter the edge of the black body aperture, even when the scan mirror is in a pristine state at the start of a deployment. Thus the heated black body, when viewed through the field stop, is probably not perfectly uniform, but may have a slightly cooler annular region around its circumference. Using the effective emissivity allows for this small effect, which depends on the geometry of the scan drum and black body aperture. The effective emissivity for both instruments was assumed in the original algorithm to be constant and was estimated by calibration experiments to have a value $\varepsilon_0 = 0.9993$. The use of this value has been confirmed by all the pre-deployment validations using the CASOTS-II external laboratory black body.

However, when the scan mirror degrades, not only does it reduce the directly reflected signal but it also scatters more stray radiation into the detector, from a wider range of angles beyond the mirror than it does when the mirror is pristine. In these circumstances, ε should not be treated as constant but must be allowed to reduce slightly, in proportion to the degradation of the scan drum mirror.

Fortunately the mirror response is readily monitored by evaluating a nominal gain factor:

$$G = \frac{C_{bb2} - C_{bb1}}{L_{BB02} - L_{BB01}} \tag{A2}$$

where C_{bb} is the detector output count when the scan mirror is pointed at a black body, the additional suffices 1 and 2 refer to the two black bodies at different temperatures, and the ideal black body radiances, L_{BB0}, are evaluated using the initial emissivity value $\varepsilon_0 = 0.9993$. That is:

$$L_{BB0} = \varepsilon_0 B_B(T_{bb}) + (1-\varepsilon_0)B_B(T_{amb}) \tag{A3}$$

G can thus be evaluated independently for every SST measurement cycle. As the mirror degrades during a deployment, G decreases as a

direct measure of the mirror degradation. Thus, assuming a linear relation between the effective emissivity and the gain we can write:

$$\varepsilon = \varepsilon_0 - f_W\left(G_{precal} - G\right) \quad \text{(A4)}$$

where G_{precal} is the gain measured during the pre-deployment validation straight after fitting a pristine scan mirror, when the degradation can be taken as zero. f_W is a constant of proportionality which may in practise be instrument dependent. For the two ISAR instruments referred to in this paper it was evaluated experimentally to be 0.45. Using Eq. (A4) the effective emissivity of each black body can be evaluated for each measurement cycle and, through Eq. (A1), is used to estimate L_{BB} and thus continuously update the internal calibration of the ISAR.

The effectiveness of this empirical adjustment to the ISAR internal calibration is confirmed as long as the post-deployment laboratory validation shows agreement between the ISAR measurement and the laboratory black body temperature to within less than 0.1 K.

Note also that the apparent gain, G, is available for use as a valuable diagnostic of the gradual degradation of the mirror surface and optical path, since it can be calculated readily from the data transmitted from ship to shore by Iridium. Should it start to vary rapidly or erratically a deployment is aborted.

References

Barton, I. J., Minnett, P. J., Maillet, K. A., Donlon, C. J., Hook, J. S., Jessup, A. T., et al. (2004). The Miami 2001 infrared radiometer calibration and intercomparison. Part II: Shipboard results. *Journal of Atmospheric and Oceanic Technology*, *21*(2), 268–283.

Barton, I. J., Prata, A. J., & Cechet, R. P. (1995). Validation of the ATSR in Australian Waters. *Journal of Atmospheric and Oceanic Technology*, *12*, 290–300.

Corlett, G. K., Barton, I. J., Donlon, C. J., Edwards, M. C., Good, S. A., Horrocks, L. A., et al. (2006). The accuracy of SST retrievals from AATSR: An initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets. *Advances in Space Research*, *37*(4), 764–769.

Corlett, G. K., & Poulter, D. J. S. (2008). An updated confidence flagging scheme for medspiration AATSR L2P files. In O. Arino, & I. S. Robinson (Eds.), *Reports from the Medspiration Project, MED-UL-REP-002-1D*. Frascati, Italy: European Space Agency ESRIN Laboratory 35 pp.

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., et al. (2002). Towards improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, *15*(4), 353–369.

Donlon, C. J., & Robinson, I. S. (1998). Radiometric validation of ERS-1 alongtrack scanning radiometer average sea surface temperature in the Atlantic ocean. *Journal of Atmospheric and Oceanic Technology*, *15*(3), 647–660.

Donlon, C. J., Robinson, I. S., Casey, K. S., Vazquez, J., Armstrong, E., Arino, O., et al. (2007). The Global Ocean Data Assimilation Experiment (GODAE) High Resolution Sea Surface Temperature Pilot Project (GHRSST-PP). *Bulletin of the American Meteorological Society*, *88*(8), 1197–1213, doi:10.1175/BAMS-88-8-1197.

Donlon, C. J., Robinson, I. S., Reynolds, M., Wimmer, W., Fisher, G., Edwards, R., et al. (2008). An infrared sea surface temperature autonomous radiometer (ISAR) for deployment aboard volunteer observing ships (VOS). *Journal of Atmospheric and Oceanic Technology*, *25*(1), 93–113.

Edwards, T., Browning, R., Delderfield, J., Lee, D. J., Lidiard, K. A., Milborrow, R. S., et al. (1990). The Along Track Scanning Radiometer measurement of sea-surface temperature from ERS-1. *Journal of British Interplanetary Society*, *43*, 160–180.

Merchant, C. J., & Harris, A. R. (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature 2. Comparison with in situ measurements. *Journal of Geophysical Research*, *104*(C10), 23579–23590.

Merchant, C. J., Harris, A. R., Murray, M. J., & Závody, A. M. (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modelling and interalgorithm comparison. *Journal of Geophysical Research*, *104*(C10), 23565–23578.

Merchant, C. J., Llewellyn-Jones, D. T., Saunders, R. W., Rayner, N., Kent, E. C., Old, C. P., et al. (2008). Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Advances in Space Research*, *41*, 1–11.

Minnett, P. J. (1991). Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *Journal of Geophysical Research*, *96*(C10), 18475–18489.

Niclòs, R., Valor, E., Caselles, V., Coll, C., & Sánchez, J. M. (2005). In situ angular measurements of thermal infrared sea surface emissivity—Validation of models. *Remote Sensing of Environment*, *94*(1), 83–93.

Noyes, E. J., Minnett, P. J., Remedios, J. J., Corlett, G. K., Good, S. A., & Llewellyn-Jones, D. T. (2006). The accuracy of the AATSR sea surface temperatures in the Caribbean. *Remote Sensing of Environment*, *101*(1), 38–51.

O'Carroll, A. G., Eyre, J. R., & Saunders, R. W. (2008). Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, *25*, 1197–1207.

O'Carroll, A. G., Watts, J. G., Horrocks, L. A., Saunders, R. W., & Rayner, N. A. (2006). Validation of the AATSR meteo product sea surface temperature. *Journal of Atmospheric and Oceanic Technology*, *23*(5), 711–726.

Parkes, I. M., Sheasby, T. N., Llewellyn-Jones, D. T., Nightingale, T. J., Závody, A. M., Mutlow, C. T., et al. (2000). The Mutsu Bay Experiment: Validation of ATSR-1 and ATSR-2 sea surface temperature. *International Journal of Remote Sensing*, *21*(18), 3445–3460.

Rice, J., Butler, J., Johnson, B., Minnett, P., Maillet, K., Nightingale, T. J., et al. (2004). The Miami 2001 Infrared Radiometer Calibration and Intercomparison. Part I: Laboratory characterization of blackbody targets. *Journal of Atmospheric and Oceanic Technology*, *21*, 258–267.

Robinson, I. S., & Donlon, C. J. (2003). Global measurement of sea surface temperature from space: Some new perspectives. *Journal of Atmospheric and Ocean Science*, *9*(1), 19–37 (previously The Global Atmosphere and Ocean System).

Robinson, I. S., Piollé J.-F., Le Borgne P., Donlon C. J., Poulter D. J. S., & Arino O. (in press). Widening the application of AATSR SST data to operational tasks through the Medspiration Service, Remote Sensing of Environment, doi:10.1016/j.rse.2010.12.019.

Wu, X., & Smith, W. L. (1997). Emissivity of rough sea surface for 8–13 μm: Modeling and verification. *Applied Optics*, *36*, 2609–2619.

Závody, A. M., Mutlow, C. T., & Llewellyn-Jones, D. T. (1995). A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer. *Journal of Geophysical Research*, *100*, 937–952.

References

Advantech (1997); *ADAM 4000 Series*; Advantech Co., Ltd.

Baldridge, A. M., Hook, S. J., Grove, C. I. and Rivera, G. (2009); The ASTER spectral library version 2.0; *Remote Sensing of Environment*; 113(4): pp. 711 – 715.

Bargeron, C. P., Hydes, D. J., Woolf, D. K., Kelly-Gerreyn, B. A. and Qurban, M. A. (2006); A regional analysis of new production on the northwest European shelf using oxygen fluxes and a ship of opportunity; *Estuarine, Coastal and Shelf Science*; Volume 69: pp. 478–490.

Barton, I. J., Prata, A. and Cechet, R. (1995); Validation of the ATSR in Australian Waters; *J. Atmos. Oceanic Technol.*; 12: pp. 290 – 300.

Barton, I. J., Minnett, P. J., Maillet, K. A., Donlon, C. J., Hook, S. J., Jessup, A. T. and Nightingale, T. J. (2004); The Miami2001 Infrared Radiometer Calibration and Intercomparison. Part II: Shipboard Results.; *J. Atmos. Oceanic Technol.*; 21: pp. 268–283.

Beggs, H. M., Verein, R., Paltoglou, G., Kippo, H. and Underwood, M. (2012); Enhancing ship of opportunity sea surface temperature observations in the Australian region; *Journal of Operational Oceanography*; 5(1): pp. 59–73.

Bell, S. (2001); *A Beginner's Guide to Uncertainty of Measurement*; National Physical Laboratory; Teddington, Middlesex, United Kingdom, TW11 0LW; 2 ed.; measurement Good Practice Guide No. 11 (Issue 2).

Berry, K. H. (1981); Emissivity of a cylindrical black-body cavity with a re-entrant cone end face.; *J. Phys. E: Sci. Instrum.*; 14: pp. 629 – 632.

Best, F. A., Revercomb, H. E., Knuteson, R. O., Tobin, D. C., Dedecker, R. G., Dirkx, T. P., Mulligan, M. P., Ciganovich, N. N. and Te, Y. (2003); Traceability of Absolute Radiometric Calibration for the Atmospheric Emitted Radiance Interferometer (AERI); in *USU/SDL CALCON*.

Bourns (2006); *4800P Series - Thick Film Surface Mounted Medium Body*; Bourns.

Castro, S. L., Wick, G. A. and Emery, W. J. (2003); Further refinements to models for the bulk-skin sea surface temperature difference; *J. Geophys. Res.*; 108(C12): p. 3377.

Challenor, P. G. (2004); Error calibration; Personal Communications.

Challenor, P. G. (2013); Error calibration - The estimation of the γ parameter; Personal Communications.

Chelton, D. B. (2005); The Impact of SST Specification on ECMWF Surface Wind Stress Fields in the Eastern Tropical Pacific; *J. Climate*; 18(4): pp. 530–550.

Corlett, G. K. (2007); Single Sensor Error Statistics; Tech. rep.; Space ConneXions Contract: 2007-12-002/C21089.

Corlett, G. K. (2012); Validation of the ATSR SST CDR and Continuity into the SLSTR Era; in *Sentinel 3 OLCI/SLSTR & MERIS/(A)ATSR workshop*.

Corlett, G. K. and Poulter, D. J. S. (2008); *An Updated Confidence Flagging Scheme for Medspiration AATSR L2P Files*; vol. MED-UL-REP-002-1D of *Reports from the Medspiration Project*; European Space Agency ESRIN Laboratory, Frascati, Italy.

Corlett, G. K., Barton, I. J., Donlon, C. J., Edwards, M. C., Good, S. A., Horrocks, L. A., Llewellyn-Jones, D. T., Merchant, C. J., Minnett, P. J., Nightingale, T. J., Noyes, E. J., O'Carroll, A. G., Remedios, J. J., Robinson, I. S., Saunders, R. W. and Watts, J. G. (2006); The accuracy of SST retrievals from the AATSR: an initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets; *Advances in Space Research*; 37: pp. 764–769.

Corlett, G. K., Merchant, C. J. and Rayner, N. A. (2012); CCI Phase 1 - Product Validation Plan; Tech. rep.; European Space Agency.

Daszykowski, M., Kaczmarek, K., Heyden, Y. V. and Walczak, B. (2007); Robust statistics in data analysis A review: Basic concepts; *Chemometrics and Intelligent Laboratory System*; 85 Issue 2: pp. 203–219.

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A. and Wood, N. (2005); A new dynamical core for the Met Office's global and regional modelling of the atmosphere; *Q.J.R. Meteorol. Soc.*; 131: pp. 1759–1782.

Donlon, C. J. and Nightingale, T. J. (2000); Effect of Atmospheric Radiance Errors in Radiometric Sea-Surface Skin Temperature Measurements; *Appl. Opt.*; 39: pp. 2387–2392.

Donlon, C. J. and Robinson, I. S. (1997); Observations of the oceanic thermal skin in the Atlantic Ocean; *J. Geophys. Res.*; 102(C8): pp. 18,585–18,606.

Donlon, C. J. and Wimmer, W. (2007); *ISAR User manual*; 1.11 ed.

Donlon, C. J., Nightingale, T. J., Fiedler, L., Fisher, G., Baldwin, D. and Robinson, I. S. (1999); The Calibration and Intercalibration of Sea-Going Infrared Radiometer Systems Using a Low Cost Blackbody Cavity; *J. Atmos. Oceanic Technol.*; 16: pp. 1183–1197.

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B. and Murray, M. J. (2002); Toward Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research; *J. Climate*; 15: pp. 353–369.

Donlon, C. J., Robinson, I. S., Casey, K. S., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Pioll, J., Barton, I. J., Beggs, H., Poulter, D. J. S., Merchant, C. J., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P. J., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R. W., Kawamura, H. and Rayner, N. A. (2007); The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project; *Bull. Amer. Meteor. Soc.*; 88: pp. 1197–1213.

Donlon, C. J., Robinson, I. S., Reynolds, M., Wimmer, W., Fisher, G., Edwards, R. and Nightingale, T. J. (2008); An Infrared Sea Surface Temperature Autonomous

Radiometer (ISAR) for Deployment aboard Volunteer Observing Ships (VOS); *J. Atmos. Oceanic Technol.*; 25: pp. 93–113.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E. and Wimmer, W. (2012); The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system; *Remote Sensing of Environment*; 116: pp. 140–158.

Donlon, C. J., Wimmer, W., Ferlet, M., Fisher, G., Nightingale, T. J. and Robinson, I. S. (2013); The Calibration and Inter-calibration of Sea-Going Infrared Radiometer Systems Using a Second-Generation Low Cost Blackbody Cavity; *In preperation.*

Eifler, W. and Donlon, C. J. (2001); Modeling the thermal surface signature of breaking waves; *J. Geophys. Res.*; 106(C11): pp. 27,163–27,185.

Embury, O., Merchant, C. J. and Corlett, G. K. (2012a); A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects; *Remote Sensing of Environment*; 116(0): pp. 62 – 78.

Embury, O., Merchant, C. J. and Filipiak, M. J. (2012b); A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Basis in radiative transfer; *Remote Sensing of Environment*; 116(0): pp. 32 – 46.

Emery, W. J., Castro, S., Wick, G. A., Schluessel, P. and Donlon, C. J. (2001); Estimating Sea Surface Temperature from Infrared Satellite and In Situ Temperature Data; *Bull. Amer. Meteor. Soc.*; 82(12): pp. 2773–2785.

ESA (2007); AATSR Product Handbook; Tech. rep.; European Space Agency.

Fairall, C. W., Bradley, E. F., Godfrey, J. S., Wick, G. A., Edson, J. B. and Young, G. S. (1996a); Cool-skin and warm-layer effects on sea surface temperature; *J. Geophys. Res.*; 101(C1): pp. 1295 – 1308.

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B. and Young, G. S. (1996b); Bulk parameterization of air-sea fluxes for Tropical Ocean-Global Atmosphere Coupled-Ocean Atmosphere Response Experiment; *J. Geophys. Res.*; 101(C2): pp. 3747 – 3764.

Fairall, C. W., Persson, P. O. G., Bradley, E. F., Payne, R. E. and Anderson, S. P. (1998); A newlook at calibration and use of Eppley Precision Infrared Radiometers. PartI: theory and application; *J. Atmos. Oceanic Technol.*; 15: pp. 1229 – 1242.

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A. and Edson, J. B. (2003); Bulk parameterization of air sea fluxes: Updates and verification for the coare algorithm; *Journal of Climate*; 16(4): pp. 571 – 591.

Fowler, J. B. (1995); A Third Generation Water Bath Based Blackbody Source; *J. Res. Natl. Inst. Stand. Technol.*; 100: pp. 591 – 599.

Fox, N. (2010); A guide to establish a Quality Indicator on a satellite sensor derived data product; Tech. rep.; A Quality Assurance framework for Earth Observation.

GCOS (2010); Implementation Plan for the Global Observing System for Climate in support for the UNFCCC; Tech. rep.; Global Climate Observing System.

Gentemann, C. L. and Minnett, P. J. (2008); Radiometric measurements of ocean surface thermal variability; *J. Geophys. Res*; 113.

Gentemann, C. L., Minnett, P. J. and Ward, B. (2009); Profiles of ocean surface heating (POSH): A new model of upper ocean diurnal warming; *J. Geophys. Res.*; 114.

GHRSST (2005); The Recommended GHRSST Data Specification (GDS); Tech. rep.; GHRSST Science Team.

Gill (2010); *WindMaster & WindMaster Pro Ultrasonic Anemometer*; Gill Instruments Limited.

Hanafin, J. A. and Minnett, P. J. (2005); Measurements of the infrared emissivity of a wind-roughened sea surface; *Appl. Opt.*; 44: pp. 398–411.

Hartman, M. C., Hydes, D. J. and Campbell, J. M. (2010); Pride of Bilbao FerryBox 2005 - an overview of the data obtained and improvements in procedures.; National Oceanography Centre Southampton Internal Document 16; National Oceanography Centre Southampton.

Hasse, L. (1963); On the cooling of the sea surface by evaporation and heat exchange; *Tellus*; XV: pp. 363–366.

Heitronics (2000); *Infrared Radiation Pyrometer KT15D*; Heitronics Infrarot Messtechnik GmbH.

Hopkins, J. (2010); Report on the analysis of outliers identified in AATSR - ISAR match-up data for 2009; Tech. rep.; University of Southampton.

JCOMM (2010a); Ship of Opportunity Programme shipping lines; http://www.jcommops.org/soopip/.

JCOMM (2010b); The Voluntary Observing Ships scheme, a framework document.; Tech. Rep. JCOMM technical report No. 4; World Meteorological Organisation.

Jessup, A. T. (2003); CIRIMS installation on the Ron Brown.; http://cirims.apl.washington.edu/Brown03Install.html.

Jessup, A. T., Zappa, C. J., Loewen, M. R. and Hesany, V. (1997); Infrared remote sensing of breaking waves; *Nature*; 385: pp. 52–55.

Jessup, A. T., Fogelberg, R. A. and Minnett, P. J. (2002); Autonomous shipboard radiometer system for in situ validation of satellite SST; in *Earth Observing Systems VII Conference, Int. Symp. Optical Sci. and Tech., SPIE, Seattle.*

Kawai, Y. and Wada, A. (2007); Diurnal sea surface temperature variation and its impact on the atmosphere and ocean: A review; *Journal of Oceanography*; 63: pp. 721–744; 10.1007/s10872-007-0063-0.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. and Saunby, M. (2011); Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties; *J. Geophys. Res.*; 116(D14): p. 14103.

Kennedy, J. J., Smith, R. O. and Rayner, N. A. (2012); Using AATSR data to assess the quality of in situ sea-surface temperature observations for climate studies; *Remote Sensing of Environment*; 116: pp. 79 – 92.

Kent, E. C. and Berry, D. I. (2005); Quantifying Random Measurement Errors in Voluntary Observing Ship Meteorological Observations; *International Journal of Climatology*; 25(7): pp. 843–856.

Kent, E. C. and Challenor, P. G. (2006); Toward Estimating Climatic Trends in SST. Part II: Random Errors; *J. Atmos. Oceanic Technol.*; 23(3): pp. 476–486.

Kipp and Zonen (2004); *CM11 Pyranometer*; Kipp and Zonen.

Kirtmann, B. and Vecchi, G. A. (2011); *The Global Monsoon System: Research and Forecast (2nd Edition)*; chap. 29. Why climate modelers should worry about atmospheric and oceanic weather, pp. 511–523; World Scientific Publishing Co Pte Ltd.

Lebigot, E. O. (2012); *Uncertainties: a Python package for calculations with uncertainties*; version 1.8.

Llewellyn-Jones, D. T., Edwards, M. C., Mutlow, C. T., Birks, A. R., Barton, I. J. and Tait, H. (2001); AATSR: Global-Change and Surface Temperature Measurements from Envisat; *ESA Bulletin*; 105: pp. 11–21.

Martin, M. J., Hines, A. and Bell, M. J. (2007); Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact; *Q.J.R. Meteorol. Soc.*; 133: pp. 981–995.

Masuda, K. (2006); Infrared sea surface emissivity including multiple reflection effect for isotropic Gaussian slope distribution model; *Remote Sensing of Environment*; 103(4): pp. 488 – 496.

Masuda, K., Takashima, T. and Takayama, Y. (1988); Emissivity of pure and sea waters for the model sea surface in the infrared window regions; *Remote Sensing Environ.*; 24: pp. 313–329.

Maxim (2008); *MAX667 - +5V/Programmable Low-Dropout Voltage Regulator*; Maxim Integrated Products.

Measurement Specialties, I. (2008); *46041 Super Stable Glass NTC Thermistor*; Measurement Specialties, Inc.

Merchant, C. J., Llewellyn-Jones, D. T., Saunders, R. W., Rayner, N. A., Kent, E. C., Old, C. P., Berry, D., Birks, A. R., Blackmore, T., Corlett, G. K., Embury, O., Jay, V. L., Kennedy, J., Mutlow, C. T., Nightingale, T. J., OCarroll, A. G., Pritchard, M. J., Remedios, J. J. and Tett, S. (2008); Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers; *Adv. Space Res*; 41(1): pp. 1–11.

Minnett, P. J. (1991); Consequences of Sea Surface Temperature Variability on the Validation and Applications of Satellite Measurements; *J. Geophys. Res.*; 96(C10): pp. 18,475–18,489.

Minnett, P. J. (2003); Radiometric measurements of the sea-surface skin temperature: the competing roles of the diurnal thermocline and the cool skin; *International Journal of Remote Sensing*; 24(24): pp. 5033–5047.

Minnett, P. J. (2004); M-AERI on the Explorer of the Seas; http://sealion.rsmas.miami.edu/explorer_maeri_test.html.

Minnett, P. J. (2011); Clarifications on SST definitions: Discussion; in *GHRSST XIII, Science Team Workshop, University of Edinburgh.*

Minnett, P. J. and Corlett, G. K. (2012); A pathway to generating Climate Data Records of sea-surface temperature from satellite measurements; *Deep Sea Research Part II: Topical Studies in Oceanography*; 77–80: pp. 44 – 51.

Minnett, P. J., Knuteson, R. O., Best, F. A., Osborne, B. J., Hanafin, J. A. and Brown, O. B. (2001); The Marine-Atmospheric Emitted Radiance Interferometer: A High-Accuracy, Seagoing Infrared Spectroradiometer; *J. Atmos. Oceanic Technol.*; 18: pp. 994–1013.

Moat, B. I., Yelland, M. J. and Molland, A. F. (2006); Quantifying the airflow distortion over merchant ships. Part II: application of the model results.; *Journal of Atmospheric and Oceanic Technology*; 23 (3): pp. 351–360.

Newman, S. M., Smith, J. A., Glew, M. D., Rogers, S. M. and Taylor, J. P. (2005); Temperature and salinity dependence of sea surface emissivity in the thermal infrared.; *Q.J.R. Meteorol. Soc.*; 131: pp. 2539–2557.

Niclos, R. I., Valor, E., Caselles, V., Coll, C. and Sanchez, J. M. (2005); In situ angular measurements of thermal infrared sea surface emissivity -Validation of models; *Remote Sensing of Environment*; 94, Issue 1: pp. 83–93.

Niclos, R. I., Caselles, V., Valor, E. and Sanchez, C. C. J. M. (2009); A simple equation for determining sea surface emissivity in the 3-15 m region; *International Journal of Remote Sensing*; 30, Number 6: pp. 1603–1619.

Nightingale, T. J. (2000); ISAR Radiance to Temperature Transformation; Personal Communications.

Nightingale, T. J. (2007); The SISTeR Validation Radiometer; http://www.atsr.rl.ac.uk/validation/sister/sis_inst/index.shtml.

O'Carroll, A. G., Eyre, J. R. and Saunders, R. W. (2008); Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations; *J. Atmos. Oceanic Technol.*; 25: pp. 1197–1207.

Ohring, G., Wielicki, B., Spencer, R., Emery, B. and Datla, R. (2005); Satellite Instrument Calibration for Measuring Global Climate Change: Report of a Workshop; *Bulletin of the American Meteorological Society*; 86(9): pp. 1303–1313.

Price, J. F., Weller, R. A. and Pinkel, R. (1986); Diurnal Cycling: Observations and Models of the Upper Ocean Response to Diurnal Heating, Cooling, and Wind Mixing; *J. Geophys. Res.*; 91(C7): pp. 8411–8427.

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T. J. and Tett, S. F. B. (2006); Improved Analyses of Changes and Uncertainties in Sea Surface Temperature Measured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset; *J. Climate*; 19(3): pp. 446–469.

Rayner, N. A., Kaplan, A., Kent, E. C., Reynolds, R. W., Brohan, P., Casey, K. S., Kennedy, J. J., Woodruff, S. D., Smith, T. M., Donlon, C. J., Breivik, L., Eastwood, S., Ishii, M. and Brandon, T. (2009); Evaluating climate variability and change from modern and historical SST observations; in Hall, J., Harrison, D. and Stammer, D. (eds.), *Proceedings of OceanObs09: Sustained Ocean Observations and Information for Society, vol. 2*; vol. 2; ESA; ESA Publ.

RBR (2004); *Temperature Logger TR-1050*; RBR Ltd.

Reid, P. C., Colebrook, J. M., Matthews, J. B. L. and Aiken, J. (2003); The Continuous Plankton Recorder: concepts and history, from Plankton Indicator to undulating recorders; *Progress in Oceanography*; 58(24): pp. 117 – 173.

Rice, J. P., Butler, J. J., Johnson, B. C., Minnett, P. J., Maillet, K. A., Nightingale, T. J., Hook, S. J., Abtahi, A., Donlon, C. J. and Barton, I. J. (2004); The Miami2001 Infrared Radiometer Calibration and Intercomparison. Part I: Laboratory Characterization of Blackbody Targets; *J. Atmos. Oceanic Technol*; 21: pp. 258–267.

Saunders, P. M. (1967); The temperature at the ocean-air interface.; *Journal of Atmospheric Sciences*; 24: pp. 269–273.

Schluessel, P., Emery, W. J., Grassl, H. and Mammen, T. (1990); On the Bulk-Skin Temperature Difference and Its Impact on Satellite Remote Sensing of Sea Surface Temperature; *J. Geophys. Res.*; 95(C8): pp. 13341–13356.

Seabird (2010); *SBE 48 Hull Temperature Sensor*; Sea-Bird Electronics, Inc.

Sexton, D. M. H., Grubb, H., Shine, K. P. and Folland, C. K. (2003); Design and Analysis of Climate Model Experiments for the Efficient Estimation of Anthropogenic Signals; *J. Climate*; 16(9): pp. 1320–1336.

Smith, T. M., Peterson, T. C., Lawrimore, J. H. and Reynolds, R. W. (2005); New surface temperature analyses for climate monitoring; *Geophys. Res. Lett.*; 32(14): p. 14712.

Soloviev, A. V. and Schluessel, P. (1994); Parameterisation of the cool skin of the ocean and of the air-ocean gas transfer on the basis of modelling surface renewals; *J. Phys. Oceanog.*; 24: pp. 1339–1346.

Stark, J. D., Donlon, C. J., Martin, M. J. and McCulloch, M. E. (2007); OSTIA : An operational, high resolution, real time, global sea surface temperature analysis system; *OCEANS 2007 - Europe conference proceedings*; pp. 1–4.

Stommel, H. (1969); Observations of the diurnal thermocline; *Deep Sea Research I*; 16: pp. 269–284.

Storkey, D., Blockley, E. W., Furner, R., Guiavarc'h, C., Lea, D., Martin, M. J., Barciela, R. M., Hines, A., Hyder, P. and Siddorn, J. R. (2010); Forecasting the ocean state using NEMO: The new FOAM system; *Journal of Operational Oceanography*; 3(1): pp. 3–15.

Stuart-Menteth, A. C. (2004); *A Global Study of Durinal warming*; Ph.D. thesis; University of Southampton.

Stuart-Menteth, A. C., Robinson, I. S. and Challenor, P. G. (2003); A global study of diurnal warming using satellite-derived sea surface temperature; *J. Geophys. Res*; 108(C5): p. 3155.

Sverdrup, H. U., Johnson, M. W. and Fleming, R. H. (1942); *The Oceans: Their Physics, Chemistry and General Biology*; Prentice-Hall, Inc.

Theocharous, E. and Fox, N. P. (2010); CEOS comparison of IR brightness temperature measurements in support of satellite validation. Part II: Laboratory comparison of the brightness temperature of blackbodies.; Tech. Rep. OP 4; National Physical Laboratory.

Theocharous, E., Usadi, E. and Fox, N. P. (2010); CEOS comparison of IR brightness temperature measurements in support of satellite validation. Part I: Laboratory and Ocean surface temperature comparison of radiation thermometers.; Tech. Rep. OP 3; National Physical Laboratory.

Tokmakian, R. and Challenor, P. G. (1999); On the joint estimation of model and satellite sea surface height anomaly errors; *Ocean Modelling*; 1: pp. 39–52.

Upper, D. (1974); The unsuccessful self-treatment of a case of writer's block; *Journal of Applied Behavior Analysis*; 7(3): p. 497.

Vaisala (1998); *HMP243 TRANSMITTER Operating Manual*; Vaisala.

Ward, B., Wanninkhof, R., Minnett, P. J. and Head, M. J. (2004); SkinDeEP: A Profiling Instrument for Upper-Decameter Sea Surface Measurements.; *J. Atmos. Oceanic Technol.*; 21: pp. 207–222.

Watts, P. D., Allen, M. R. and Nightingale, T. J. (1996); Wind speed effects on sea surface emission and reflection for the Along Track Scanning Radiometer; *Journal o fAtmospheric and Oceanic Technology,*; 13(1): pp. 126–141.

Webster, P. J., Clayson, C. A. and Curry, J. A. (1996); Clouds, Radiation, and the Diurnal Cycle of Sea Surface Temperature in the Tropical Western Pacific; *J. Climate*; 9(8): pp. 1712–1730.

Wimmer, W. and Robinson, I. S. (2008a); Report on WP 700: A New Method for Quality Assessment of SST Validation Match-ups; Tech. rep.; National Oceanography Centre, Southampton.

Wimmer, W. and Robinson, I. S. (2008b); Validation of AATSR Sea Surface Temperature Products using the ship borne ISAR Radiometer - Phase 2 Final Report; Tech. rep.; National Oceanography Centre, Southampton.

Wimmer, W. and Robinson, I. S. (2008c); WP800 report; Tech. rep.; National Oceanography Centre, Southampton.

Wimmer, W., Robinson, I. S. and Donlon, C. J. (2012); Long-term validation of AATSR SST data products using shipborne radiometry in the Bay of Biscay and English Channel; *Remote Sensing of Environment*; 116: pp. 17–31.

Wu, X. and Smith, W. L. (1997); Emissivity of rough sea surface for 813 m: modeling and verification,; *Appl. Opt.*; 36: pp. 2609–2619.