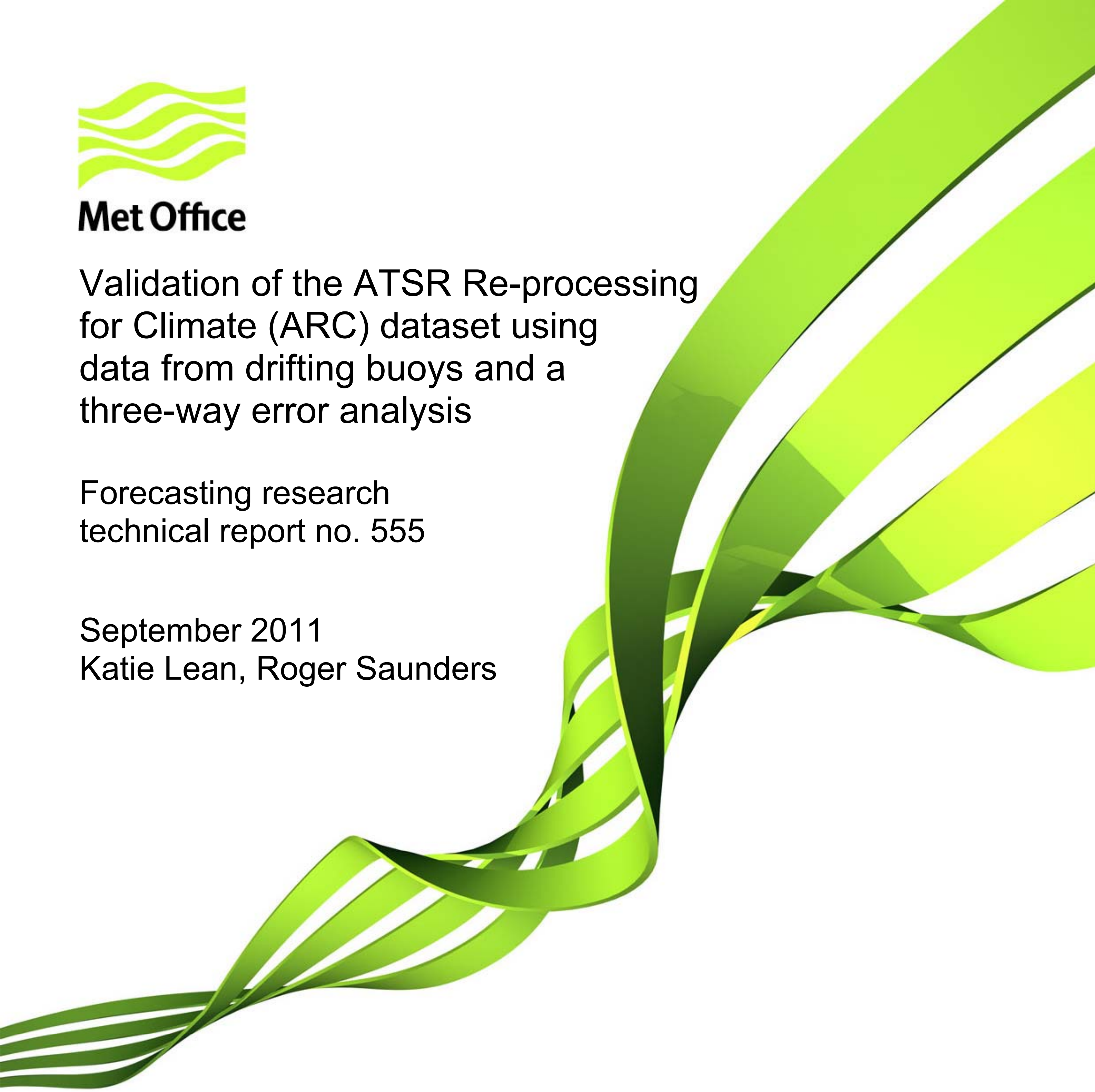
Met Office

Validation of the ATSR Re-processing for Climate (ARC) dataset using data from drifting buoys and a three-way error analysis

Forecasting research
technical report no. 555

September 2011
Katie Lean, Roger Saunders



Abstract

The Along Track Scanning Radiometer (ATSR) Reprocessing for Climate (ARC) project aims to create a data set of sea surface temperatures (SSTs) covering recent decades that can be used for the first independent check on climate change analysis. Here, the ARC SSTs are validated using comparisons with collocated drifting buoy observations and using a three-way error analysis which also includes Advanced Microwave Scanning Radiometer (AMSR-E) data. During the analysis, the SSTs using the three channel night-time retrievals in the ARC data are found to have a warm bias of 0.054K (standard deviation 0.151K) with respect to the drifting buoy data for the 1995-2009 time period using ATSR-2 and AATSR instrument data. However, when studying the two channel retrievals a noticeable difference is found between the ATSR-1 data and later years. Some dependence on latitude, season and fields such as total column water vapour are also found in the ATSR-2 and Advanced Along Track Scanning Radiometer (AATSR) period (the data for ATSR-1 did not disagree with these results but larger confidence intervals often meant that trends were not significant). The three-way error analysis revealed a small standard deviation of error of 0.14K for the ARC bulk SSTs using the three channel night-time retrieval.

1. Introduction

Accurately assessing the trend in sea surface temperature (SST) is very important for climate change studies – changes in the processes by which the oceans store and transport heat may have significant impacts on the climate. In order to better understand any variation in the SST, monitoring is needed on a global scale and over a reasonable time period covering recent years where climate change has been most rapid.

Traditionally buoy and ship data have been used for looking at SST trends. However, there are several disadvantages to using this data such as variation in the depth of the measurement and the inadequate maintenance of the buoys or poor calibration at launch which leads to some inaccurate readings. Retrieval of SSTs from instruments onboard satellites provides much better global coverage and has the potential to give much more accurate, well calibrated results. The ATSR Re-processing for Climate (ARC) project aims to develop an improved, highly accurate data set suitable for use in climate change analyses and in the study of oceanographic processes such as the progress of the Western Boundary currents (Llewellyn-Jones, 2006). Further information about the ARC project can be found in Merchant, *et. al*, 2007. If successful, this project will be able to provide the first independent check on estimates of the rate of warming of the ocean surface covering recent decades.

In order to verify the accuracy of the ARC data, validation was carried out using drifting buoy data. SSTs from the two data sources were compared using a co-location technique where a buoy observation is matched to a satellite reading if certain criteria are met. During the analysis, regional, global and latitudinal biases were investigated as well as considering biases as a function of wind speed and insolation through using data from the ERA-Interim reanalysis produced by ECMWF. Further to this, a three way error analysis was carried out which, with the inclusion of a third SST data source, allows the estimation of the standard deviation of the error in the ARC data. For this part of the study, SSTs from the Advanced Microwave Scanning Radiometer – Earth observing system (AMSR-E) were also used.

2. Data sets

2.1 ARC dataset

The Along Track Scanning Radiometer (ATSR) instruments were built with a specific aim to provide an accurate SST record. The first instrument, ATSR-1, was launched on the European Remote Sensing satellite (ERS-1) in July 1991 in a sun synchronous orbit. Subsequently, ATSR-2 and the Advanced ATSR (AATSR) were launched on ERS-2 in April 1995 and on ENVISAT in March 2002 respectively, also in sun synchronous orbits. All the instruments were very well calibrated and this has led to a continuous and accurate record of SST spanning two decades. Data provided for this project extended from August 1991 to December 2009. ARC data using the AATSR instrument were accessible from late July in 2002 although files using ATSR-2 data were available until mid 2003. In order to maximise the use of data derived from AATSR, during the analysis in this report, observations processed from ATSR-2 have only been used until 23/07/2002. In 1995 and 1996, ATSR-2 data is used where possible in preference to ATSR-1. Rare problems with the instruments meant that some days were unavailable such as for a period of almost three weeks in early 2001 when ATSR-2 shutdown due to a gyro failure on the spacecraft.

The ATSR instruments are dual view radiometers with one aperture directed in the nadir view and the second, forward view at a viewing angle of 55° to the zenith. When a SST retrieval is made, there is a choice to use only the nadir view or both views and either two instrument channels (10.8μm and 12μm) or three channels (10.8μm, 12μm and 3.7μm). The three channel retrieval is only valid at night-time due to sunlight contamination of the 3.7μm channel during the day. The instruments were intended to give a dual view retrieval of the SST to an accuracy of 0.3K while seeking a long term stability of 0.1K/decade. Throughout this report, all SSTs found through three channel retrievals will only be from night-time observations and the dual view was used for both two and three channel retrievals.

While the (A)ATSR data is of a good quality, corrections to the data have been needed and the ARC project aims to make further improvements, for example better methods of cloud detection and further analysis of the instrument overlap periods in order to correct instrumental drift (Llewellyn-Jones *et. al*, 2007). This report aims to assess whether these changes have produced a more accurate data set. Throughout the analysis presented here, ARC data using the full Bayesian cloud mask – with the most reliable detection – were used apart from two channel retrievals during the ATSR-1 period. The 3.7μm channel failed on ATSR-1 in May 1992. This meant that three channel retrievals were only available until this time and two channel retrievals at night were not available when using the Bayesian full cloud mask from May 1992. However, for the day time data either the full Bayesian or minimum Bayesian cloud mask can be used throughout the ATSR-1 period. Before the failure of this channel, there was a problem in switching between day and night modes so prior to this there are very few day time retrievals.

The ARC data set was supplied in daily files and the time of the observations was given in seconds after midnight. The local equator crossing time for AATSR was 30 minutes earlier than for ATSR-2 and ATSR-1. To avoid any discontinuities in the long term record, the SSTs at depth for AATSR were calculated to be correct for 30 minutes after the observation time. However, the affected records in the ARC data were still provided with the original observation time which corresponds to the skin SST retrieved. In order to carry out the validation with the buoys, the 30 minute correction was applied to the observation time before the collocation process was carried out.

The data contained the two and three channel retrieved skin SSTs and the corrections needed to produce the sub-skin temperature and the temperature at depths of 0.2m, 1m and 1.5m. The 1m measurement is taken here to typically represent the depth at which the buoys measure. Several other fields for each grid point were included such as the total column water vapour and the uncertainty of the SST retrieval which is a combination of the theoretical performance of the retrieval, number of pixels present and the variability in each cell. Further details regarding the contents of the ARC files can be found in Embury, 2011. All SSTs and accompanying fields were presented on a 0.1° x 0.1° grid. Values of solar flux and wind speed for each cell were acquired from the ERA-Interim Reanalysis.

2.2 Buoy data

Two different sources of buoy data were used for the collocation process and only drifting buoys were selected. The number of drifting buoy observations per day increases quite dramatically throughout the period studied. From 1991-1996, the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) version 2.5 was used. The ICOADS contains data acquired over a time scale of centuries and includes drifting and moored buoys, ship and platform observations. The location of a buoy observation in the ICOADS files used here was only recorded to an accuracy of 0.1° making collocations less accurate.

From 1997 onwards, drifting buoy data from the Global Telecommunications System (GTS) were used. More observations can be found in the GTS files in comparison to the ICOADS files and locations of measurements are given to three decimal places.

2.3 AMSR-E data

AMSR-E data are available from June 2002 and were used in this validation project as a third source of data in order to carry out the three way error analysis. AMSR-E was launched in May 2002 on NASA’s AQUA satellite and is used to retrieve information about SSTs, atmospheric water vapour, cloud vapour and rain rate at microwave wavelengths allowing the instrument to view the surface during cloudy conditions. Version 5 data were obtained in daily averaged files and fields are given at a resolution of 0.25°. The SSTs correspond to an ocean depth of a few millimetres.

3. Method

3.1 ARC data and buoy collocation process

The Met Office currently monitors AATSR SSTs through comparison with collocated drifting and moored buoys with the analysis being updated daily with the most recently received data. Code used for this process was adapted to use the ARC data and to add relevant model fields from the ERA-Interim reanalysis.

3.1.1 Initial quality control

Before the match-up process was carried out, the buoy and ARC data were initially subject to quality control. For all years, the unique identification numbers of the buoys were checked against a list of known unreliable buoys and areas of ocean prone to inaccurate readings and rejected if on the list. The buoy data also contained a quality control flag from previous tests carried out – these considered, for example, if the buoy SST was more than 8K from the climatology value of the area or if parameters such as the longitude and latitude were valid. For the years 2002-2009, the buoy records also included corresponding NWP model predictions of SST for the same location and time. For these years, the matchup-up process was run on weekly batches of data and considered the differences between the measured SST and model SST for records from the same buoy throughout the week period. The mean and standard deviation of these differences were calculated and compared to thresholds so that if a buoy displayed too much variation within the week or showed a considerable bias then all the records from that buoy would be rejected under the assumption that it was unreliable.

The ARC data contained a quality control flag which was used to remove any records where no correction was applied to adjust the skin temperature to the temperature at various depths. It was also found that some potential thermoclines remained in the daytime data (although the number was very small in comparison to the total number of match-ups later produced). A potential thermocline was flagged and the record removed where the ARC sub-skin temperature was predicted to be more than 0.2K warmer than the bulk temperature.

3.1.2 Collocation criteria

The ARC data were given on a grid of 0.1° square cells so that the latitude and longitude of the cell boundaries were given to one decimal place. This meant that the buoys in the ICOADS, which were given to a precision of 0.1°, were located on the boundary of four cells. In the matchup process, these four cells were considered for each buoy and should more than one provide a successful collocation, the one with the shortest time difference between the measurements was chosen. For the buoys in the GTS data set, the three decimal place precision meant that a single corresponding cell could be found. In the case that a buoy might fall on the cell boundary the cell at the lower latitude and longitude position was selected to check for collocation.

If a buoy SST was found in the same cell as a satellite measurement then, to be successful, the time difference between the two observations was required to be less than 180 minutes.

Only one successful collocation was allowed for each buoy so in the event that it matched with more than one satellite measurement, e.g. from multiple overpasses of the satellite, then the one with the shortest time difference was selected. Duplicate match-ups, where the same buoy produces more than one measurement which matches to the same satellite observation were also screened by taking the records with the smallest time difference again.

Collocations where there was any ice recorded in the field of view of the satellite were also discarded. The ARC data were supplied with the SST model predictions from ECMWF so match-ups were successful if the buoy measurement was within ±5K of the model estimate (this extra check could be more beneficial for buoys from before 2002 where less initial quality control could be carried out).

When carrying out the collocation process the sensitivity to the match-up criteria was investigated. It was found that when duplicates were included and buoy data flagged as unreliable in the initial quality control were allowed to pass, there was still very little change in overall mean bias for the whole time period despite the rise in the number of matchups. The standard deviation was observed to increase which corresponded to the larger amount of poorer data. The use of different time windows and changing the allowed distance between observations was considered further when carrying out the three-way error analysis.

3.2 Analysis of the match-ups

3.2.1 Statistical analysis

The quality of the ARC data was investigated in a number of ways. Global and regional statistics concerning the ARC – buoy bias were considered. This included looking at weekly average biases as well as longer, yearly statistics. The zonal trend and seasonal variations were also studied. Histograms were constructed to gain a better understanding of the spread of the biases and comparisons between match-ups from the different instruments were made.

Further to this, an investigation was also carried out to look at the biases as a function of wind-speed, insolation, total column water vapour, the time difference between the collocated measurements and various other fields.

In order to account for the spatial variation when considering the global statistics, biases were collected into 1° x 1° cells and measurements within each box were averaged. This helped to ensure that cells towards the high latitudes containing fewer match-ups were not obscured by the higher densities in lower latitudes. Using a simple cosine, area weights were calculated for each cell which were related to the area of the box on the surface of the Earth. This gives each cell an equal contribution to the overall mean so that grid boxes with larger areas on the globes do not skew the mean. Before collecting the data onto the 1° grid, a mean and standard deviation were calculated from the raw differences. Those match-ups exceeding ±3 times the standard deviation from the mean were rejected and successful collocations were then averaged in the 1° grid boxes.

The areas of ocean selected for the regional statistics are listed in Table 1. Since a smaller latitude range of the globe is selected, a simpler method of calculating the mean and standard deviation was used which involved the exclusion of biases more than ±3 times the standard deviation from the mean before recalculating the statistics. This method of two passes to find the mean and standard deviation should help to remove outliers and was employed wherever possible before further processing in other aspects of the graphical analysis such as the zonal statistics.

Ocean region	Latitude min	Latitude max	Longitude min	Longitude max
Tropical Atlantic	-20	20	-40	0
North Atlantic	40	60	-40	-15
Southern Ocean	-90	-40	-180	180
Indian Ocean	-30	0	60	90
West Pacific	-15	30	140	180
East Pacific	-30	30	-180	-120

Table 1 Latitude and longitude limits for ocean basins used for regional bias statistics

Zonal means were at first calculated using 1° bands but the resulting graphs revealed a lot of noise while in the polar regions means based on the very small numbers of collocations available produced some extreme peaks in the graphs. In order to avoid the rapid variation but also retain larger scale trends, the zonal means were calculated using 3° bands. However, for the ATSR-1 data it was necessary to use 6° bands due to the small number of collocations.

Error bars on the graphs represent the 95% confidence limits calculated from the standard error. Values were taken from Student's t-distribution (Knight, 2010) to multiply the standard error to the required confidence level. The value for *t* was allowed to decrease for up to and including 200 degrees of freedom and for larger data numbers took the value of 1.96 which applies as the number of degrees of freedom tends to infinity.

3.2.1 Three-way error analysis

The method of the three-way error analysis enables the calculation of the standard deviation of error on each observation type. Work carried out by O'Carroll *et. al*, 2008, using AATSR three channel night-time retrievals from 2003 (and similarly comparing with AMSR-E and drifting and moored buoy SSTs) found that the error variance can be calculated by the following equation:

$$\sigma_x^2 = 0.5 * (V_{xy} + V_{zx} - V_{yz}) \qquad \textbf{(1)}$$

where σ_x^2 is the variance of the error in observation type *x* and V_{xy} is the variance between two observation types, *x* and *y*. The derivation and discussion of this result can be found in O'Carroll *et. al*, 2008. The report found that when using drifting buoys only with a three hour time window, the standard deviation of error for the AATSR bulk three channel retrieval SST was 0.14K, the standard deviation of error for buoy SST was 0.24K and for AMSR-E SST was 0.42K. The AMSR-E data has 0.25° resolution so the nearest grid point to the location of the ARC data in each of the match-ups was chosen. The standard deviation of the differences between each combination of two data sources was calculated and substituted in equation (1).

In equation (1) there is no need to define the temporal or spatial scales being used. The method relies on using scales for which the covariances of the errors of representativeness (errors concerning the "difference between the value of the variable on the space/time scale on which it is actually measured and its value on the space/time scale on which we wish to analyse it" (O'Carroll *et. al*, 2008)) are negligible compared to the error covariances. This assumption allows simplification to the equation above. In order to determine whether this is valid in the case of the data used here, five different experiments were carried out which vary the criteria by which the ARC and buoy data were collocated (Table 2). The relaxed spatial criterion in experiment 4 was chosen to investigate the impact of the lower resolution of the ICOADS. There is also a brief discussion of the sensitivity of the statistics to the match-up criteria. The ARC three channel night-time retrievals are used for all the experiments.

Experiment no.	Criteria
1	Buoys collocated to same cell as ARC observation, time window of 180mins
2	Buoys collocated to same cell as ARC observation, time window of 60mins
3	Buoys collocated to same cell as ARC observation, time window of 240mins
4	Buoys collocated to within 0.1° of ARC observation, time window of 180mins
5	Buoys collocated to within 1° of ARC observation, time window of 180mins

Table 2 Details of the different match-up criteria to test the three-way error analysis method

4. Results

4.1 Statistical analysis of the match-ups

4.1.1 Global statistics

The number of collocations used in the global statistics for each year (Figure 1) rises very steeply after around 1997. As 1995 and 1996 were incomplete years for the three channel retrieval, the total number of match-ups appear low compared to the two channel retrievals. The total for 1991 is also small as the data were only available from August. The dip in numbers in 2001 may be partly due to the shutdown of ATSR-2 for nearly three weeks around the end of January however the general rise in the number of match-ups is a reflection of the increasing number of drifting buoys. The ARC data generally contained more day than night-time data leading to more successful match-ups during the daytime.

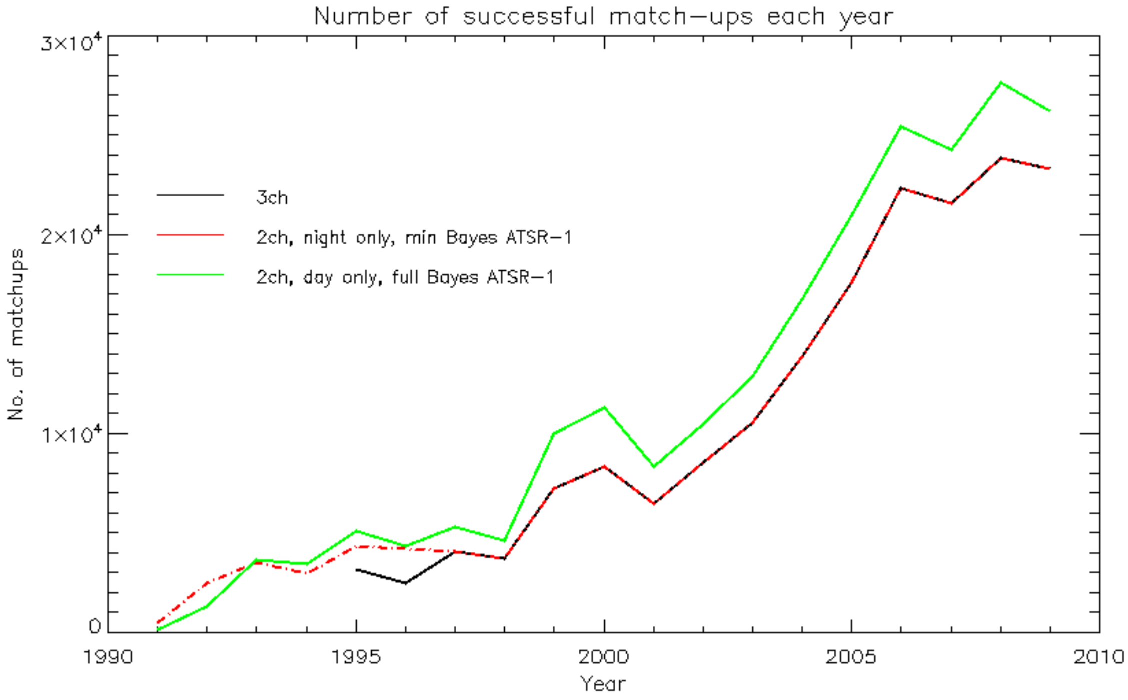


Figure 1 Graph showing the number of successful match-ups used in the calculation of global statistics

Figure 2 shows the global distribution of the biases on a 1° resolution grid for the three channel retrievals throughout the time period. The mean (as calculated by the method outlined in section 3.2.1) is 0.054K, with a standard deviation of 0.151K. There is good global coverage and the mean is relatively low although the later years will dominate the statistics.

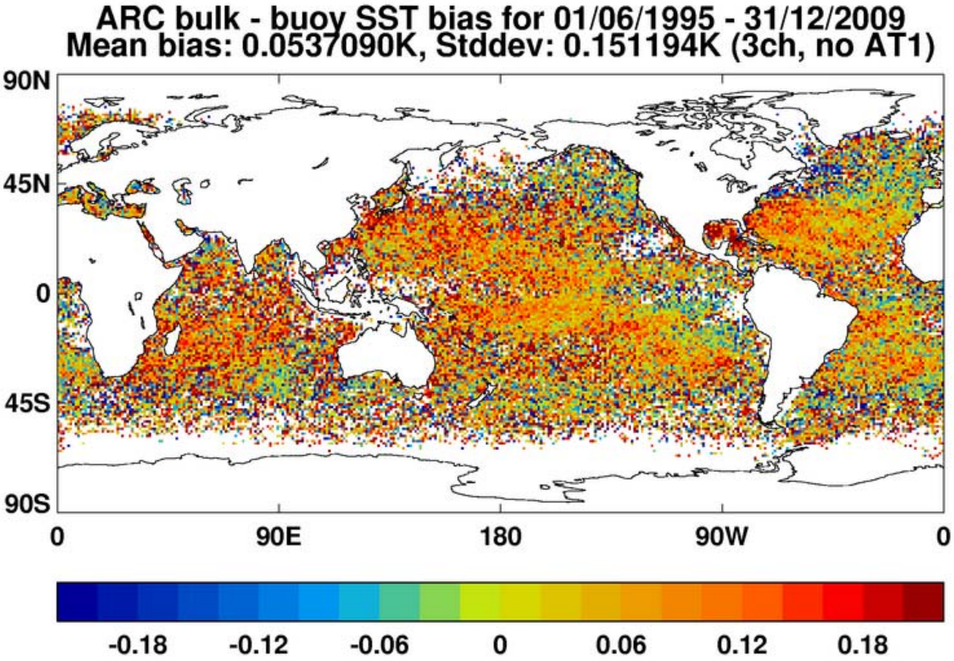

ARC bulk - buoy SST bias for 01/06/1995 - 31/12/2009
Mean bias: 0.0537090K, Stddev: 0.151194K (3ch, no AT1)
90N
45N
0
45S
90S
0
90E
180
90W
0
-0.18
-0.12
-0.06
0
0.06
0.12
0.18

Figure 2 Global map of ARC bulk – buoy biases for 3 channel night-time retrievals for 1995-2009 (no ATSR-1 data) using 1° resolution grid.

The yearly global means are shown in Figure 3 where for the ATSR-1 period, the data using the full Bayesian cloud mask and the minimum Bayesian cloud mask are shown separately. The AATSR period is the most stable for all the types of the retrievals with the two channel night-time retrieval consistently producing slightly lower mean values than for the three channel retrieval.

The time period for the ATSR-2 instrument (mid 1995 – mid 2002) shows slightly larger variation in the global mean bias although the transition between years is relatively smooth. The fall in bias during 1997-1998 could be attributed to the strong El Niño occurring. It may also be a consequence of a high proportion of the match-ups being located in the higher latitude regions where the satellite can sometimes produce lower biases. The ocean currents around the equator tend to act to sweep the buoys northwards and southwards leading to lower concentrations of collocations in the low latitude regions. Later years tend to have more global coverage and higher densities of collocations reflecting the far larger number of drifting buoys.

The yearly means for ATSR-1 vary much more between years and show some reasonably large biases in the two channel night-time compared to AATSR. There is also a larger difference between the two channel night-time and daytime data. The difference between using the full or minimum Bayesian cloud masks in the daytime data is not very significant.

The reduction in standard deviation of the bias seen in the later years (Figure 4) is partly correlated with the dramatic increase in the number of match-ups seen in Figure 1 which also leads to narrower 95% confidence limits. The greater variability seen in 1991-1996 could be explained by the smaller number of match-ups but the large standard deviation also shows that there is generally a tendency towards more extreme values (this is discussed further below when considering Figure 5). The use of the full Bayesian cloud mask appears to produce slightly lower standard deviations, apart from in during 1991, for the day time collocations. This supports the concept that the full cloud mask should be more effective in cloud detection and would therefore reduce the number of cloud contaminated observations. Appendix A considers the difference in the full Bayesian and SADIST cloud masks for 2005-2009 where many more collocations are available.

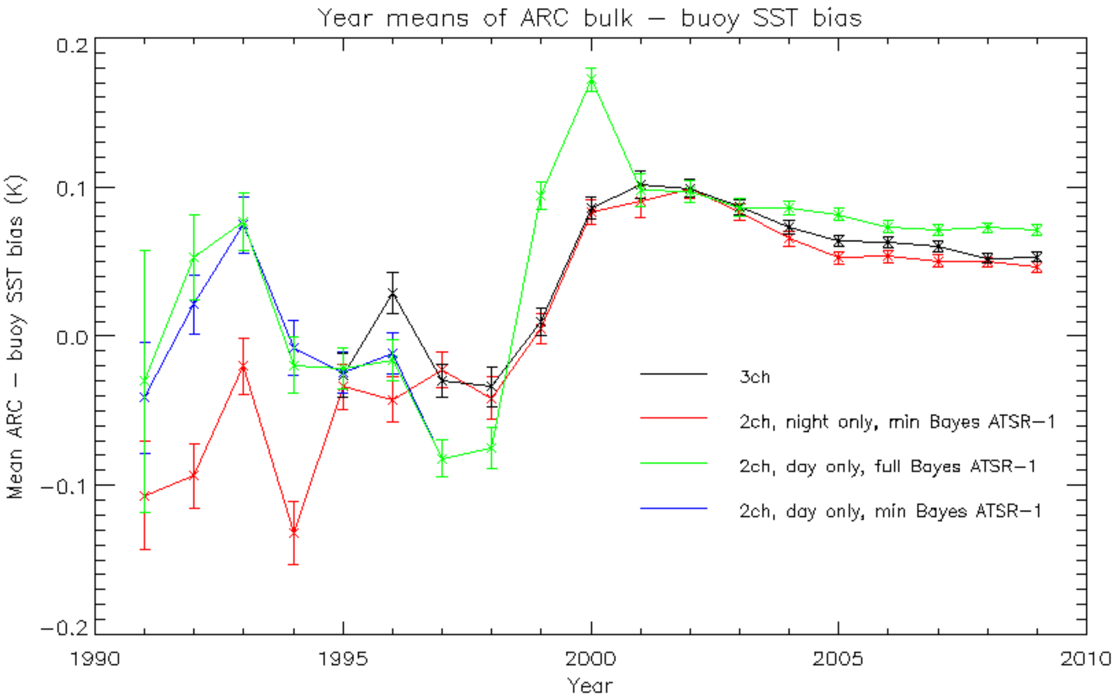


Figure 3 Year global means of the ARC bulk – buoy SST bias with 95% confidence limit (in calculating the global means, a simple mean and standard deviation were calculated and biases exceeding three standard deviations from the mean were rejected. Remaining biases were placed in a grid of 1° x 1° cells and individual cells were averaged before area weights were applied and an overall global average was calculated)

The histograms in Figure 5 give a clearer idea of the variation of the biases and support the trend of standard deviations discussed earlier. The graph for 1993 reveals the large spread around the mean value and the higher proportion of extreme values (both warm and cold biases). As the years progress, a much higher percentage of readings can be found close to the mean and the number of biases further from the mean falls off more quickly. Similar magnitudes and spreads are also observed in the daytime biases as well.

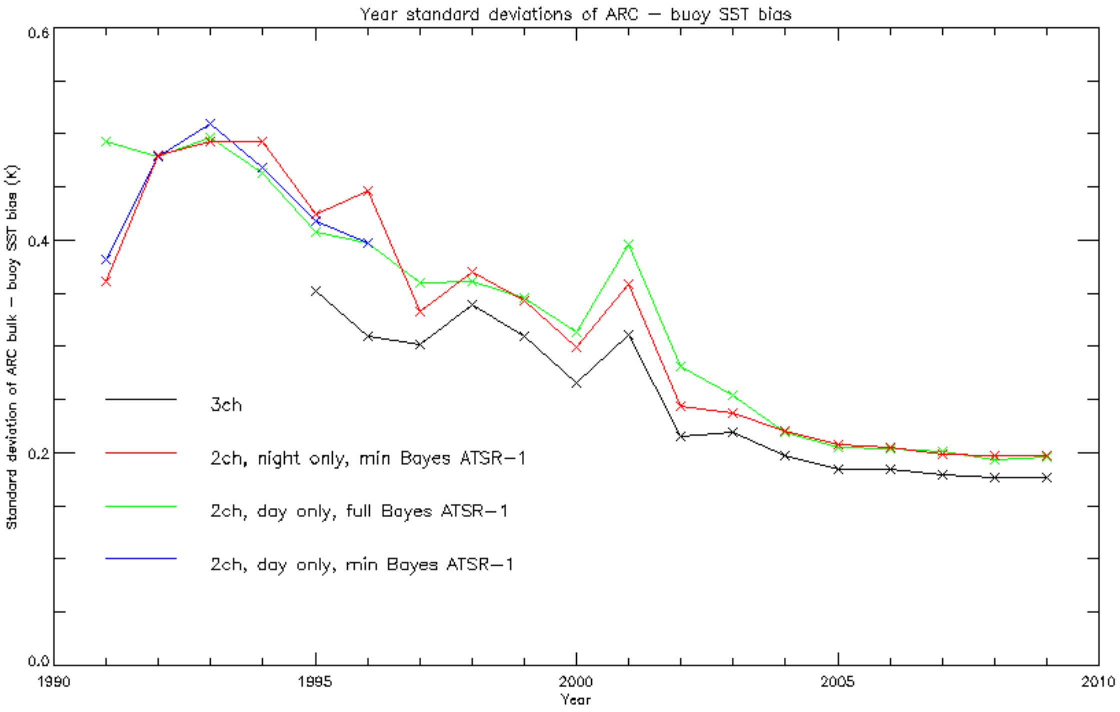


Figure 4 Year standard deviations of ARC bulk – buoy SST bias

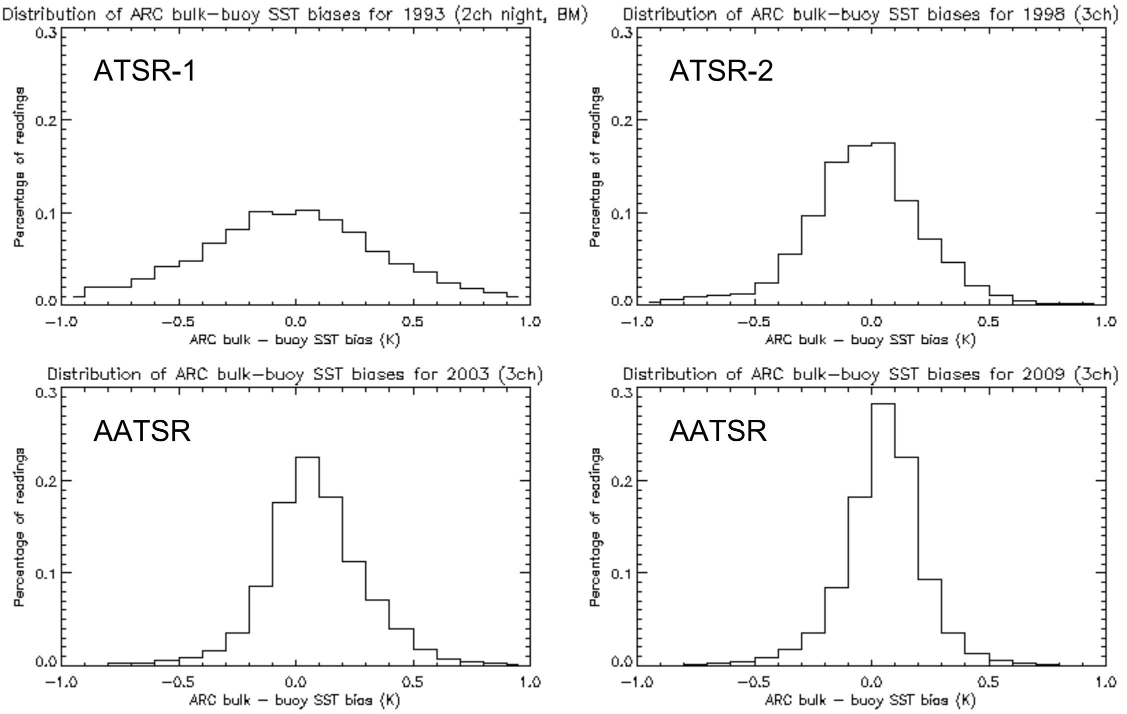


Figure 5 Histograms showing the distribution of biases globally for 1° grid boxes for a selection of years using the three channel retrieval SSTs (bin size = 0.1K)

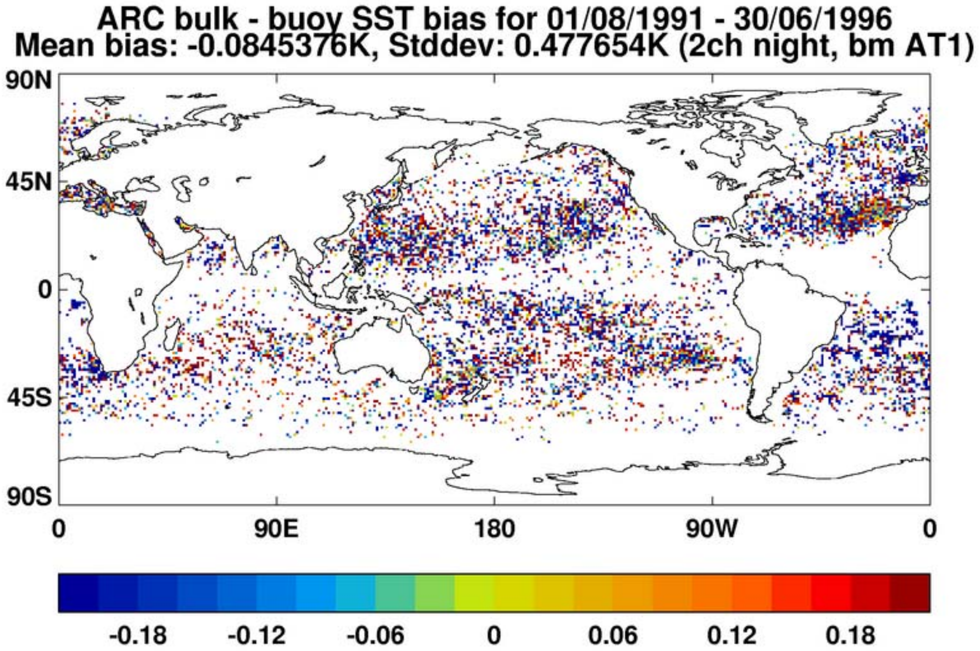
4.1.2 Instrument comparison

A summary of the mean global biases and standard deviations for the different instrument periods is presented in Table 3. The standard deviations decrease as the instrument becomes more recent indicating that although the mean bias is quite low for the ATSR-1 daytime two channel SSTs there is also a lot of variability around this value (as highlighted in Figure 5). There is also quite a large difference in the overall mean bias for the day and night-time in ATSR-1 as observed in Figure 3. The statistics of the whole time series also demonstrate how the later years dominate and mask some of the extreme biases observed in the early years. All the mean biases for the different instruments are found to be less than a magnitude of 0.1K although as Figure 5 shows, there is still an element, particularly in the earlier years, of balance of extreme warm and cold biases rather than collocations with a consistently small bias.

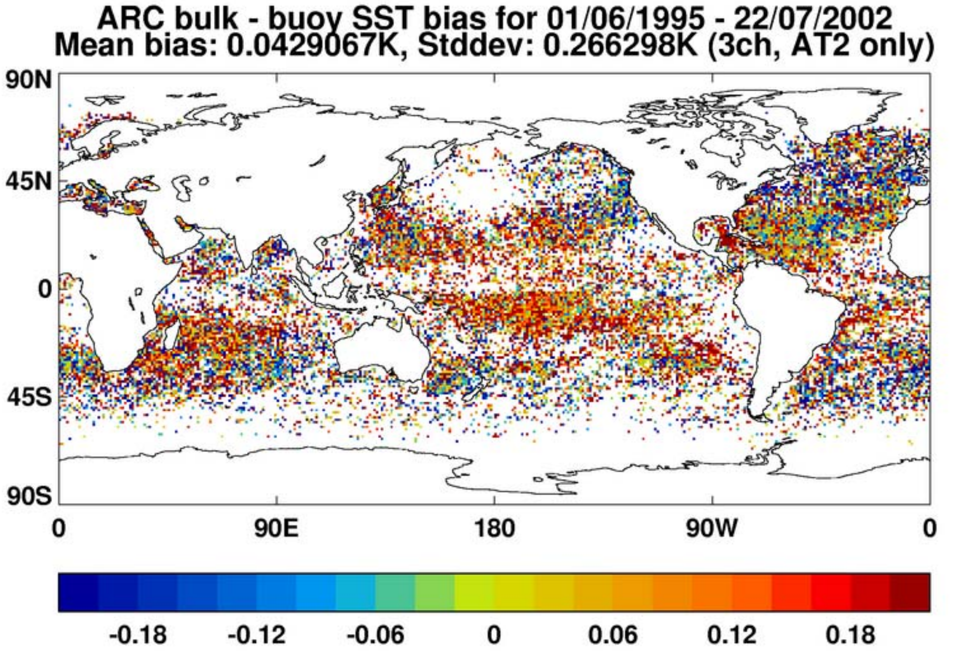
Channel selection	ATSR-1 bm = Bayes min mask bf = Bayes full mask		ATSR-2		AATSR		Whole time series 2ch: 1991-2009 3ch: 1995-2009	
	Mean (K)	Std. dev (K)	Mean (K)	Std. dev (K)	Mean (K)	Std. dev (K)	Mean (K)	Std. dev (K)
3 ch night only	-	-	0.043	0.266	0.059	0.143	0.054	0.151
2 ch night only	-0.085 (bm)	0.478 (bm)	0.040	0.296	0.053	0.159	0.044	0.182
2 ch day only	0.008 (bm) 0.008 (bf)	0.470 (bm) 0.471 (bf)	0.066	0.321	0.072	0.158	0.064	0.184

Table 3 Summary of mean global biases and standard deviations using 1° grid boxes for the different instrument periods and the complete time series

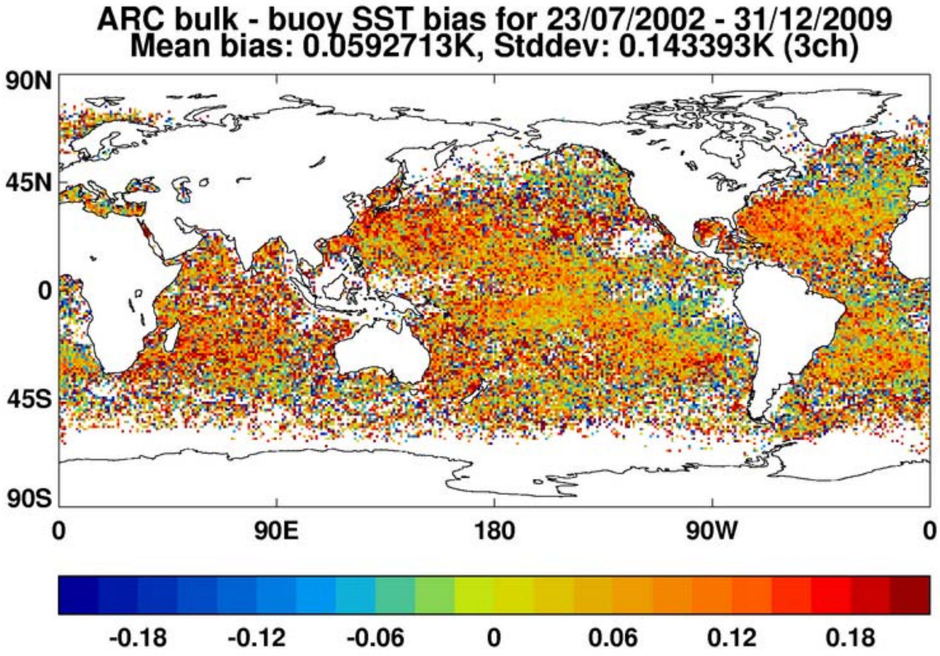
Figure 6 (a)-(c) look at the difference in the global distribution of the biases using the different instruments. The time period for ATSR-2 contained only about a quarter of the number of collocations for AATSR so even though it spans over five complete years the global coverage is still not as complete. Match-ups from the earlier period are also concentrated towards the mid and low latitudes which tend to have warmer biases (this is discussed further in sections 4.1.4 and 4.1.5). The more extreme nature of the biases during the ATSR-1 period is illustrated again.


ARC bulk - buoy SST bias for 01/08/1991 - 30/06/1996
Mean bias: -0.0845376K, Stddev: 0.477654K (2ch night, bm AT1)
90N
45N
0
45S
90S
0
90E
180
90W
0
-0.18
-0.12
-0.06
0
0.06
0.12
0.18

(a)


ARC bulk - buoy SST bias for 01/06/1995 - 22/07/2002
Mean bias: 0.0429067K, Stddev: 0.266298K (3ch, AT2 only)
90N
45N
0
45S
90S
0
90E
180
90W
0
-0.18
-0.12
-0.06
0
0.06
0.12
0.18

(b)


ARC bulk - buoy SST bias for 23/07/2002 - 31/12/2009
Mean bias: 0.0592713K, Stddev: 0.143393K (3ch)
90N
45N
0
45S
90S
0
90E
180
90W
0
-0.18
-0.12
-0.06
0
0.06
0.12
0.18

(c)

Figure 6 (a)-(c) Global bias maps comparing ATSR-1 (01/08/1991 – 30/06/1996), ATSR-2 (01/06/1995 – 22/07/2002) and AATSR (23/07/2002 – 31/12/2008) using the two channel night time retrieval with minimum Bayesian cloud mask for ATSR-1 and the three channel retrieval otherwise

4.1.3 Weekly mean biases and seasonal variation

Figure 7 shows the global weekly mean biases for a sample of years. It again reveals the large rise in the number of successful collocations in the later years as reflected by the greater range of the confidence intervals in 1993 and 1998 compared to 2003 and 2009. As the years progress, the range in the value of the week mean bias becomes smaller and the weekly standard deviations (not shown) also decrease and become more stable. In agreement with Figure 3, the weekly means from the year 2000 onwards tend to be slowly decreasing. The lower biases seen in the yearly means in 1997 and 1998 were also observed in the weekly values and appear to continue until about September 1999 before rising again.

The graphs of the weekly means also show how the night-time two channel retrieval is not significantly different from using the three channel retrieval at night-time when considering the 95% confidence intervals. The two channel retrieval at daytime produced similar patterns in weekly means to the three channel retrieval but frequently the bias value was slightly higher apart from during the 1995-1998 period.

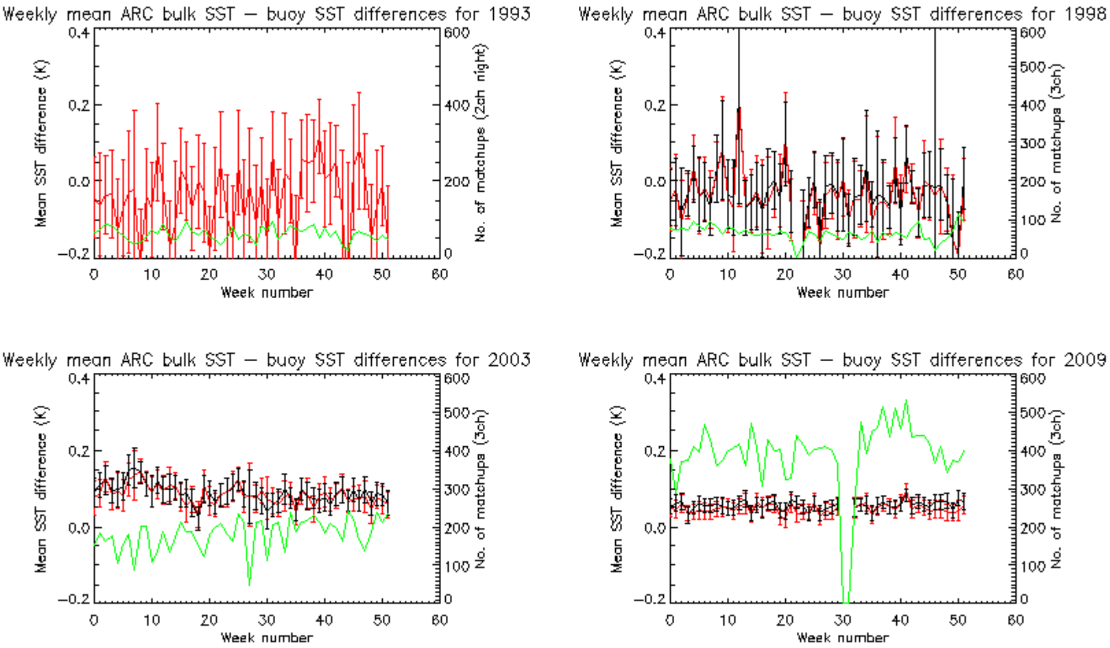


Figure 7 Selection of graphs showing global weekly means of ARC bulk – buoy SST bias. In black: 3 channel retrieval with 95% confidence intervals marked; red: 2 channel night-time retrieval with 95% confidence intervals marked (minimum Bayesian cloud mask used for 1993); green: number of match-ups per week for the 3 channel retrieval.

The seasonal variation in the three channel ARC bulk – buoy SST bias is shown in Figure 8 where the SST differences for the same week of each year were averaged over the years 1997-2009 for different latitude bands of the globe. Data from before 1997 were not included and instead treated separately in order to investigate the more stable part of the time series independently. The Northern hemisphere (20°N - 90°N) and Southern hemisphere (90°S - 20°S) both show the bias decreasing during the summer months. In the tropics there is no obvious seasonal trend with the mean bias appearing slightly higher than the global average for the time period for most weeks. These higher biases near the equator may be due to the increased cloud cover causing more error in the retrievals.

The seasonal variation for the ATSR-1 period shows the means covering a slightly larger range of values – the patterns supporting the idea of higher biases during the winter months. However, due to the large confidence intervals, the trend does not appear very significant for the 95% confidence level. The tropics, as in the case of the later years, do not show any seasonal pattern.

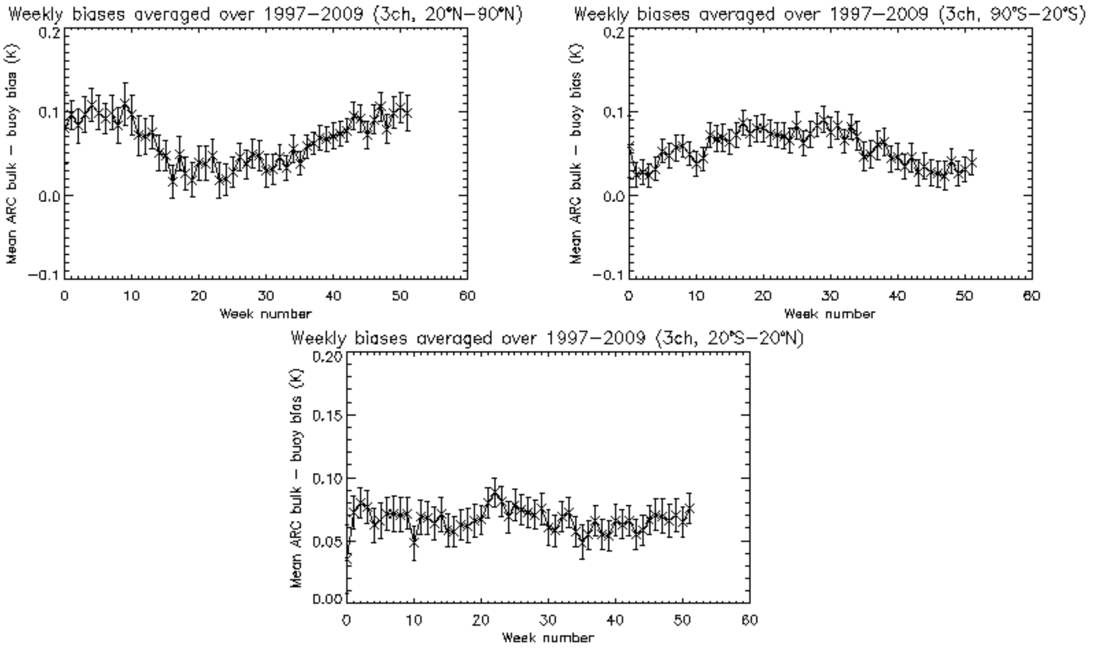


Figure 8 Graphs showing the global weekly mean biases averaged over 1997-2009 for the Northern hemisphere, Southern hemisphere and tropics using 3 channel night-time retrievals

4.1.4 Zonal variation of biases

Plots of the zonal mean biases using latitude bands of 3° (Figure 9) show slightly higher biases are observed in the tropical regions and a larger range of the 95% confidence limits are typical in the high latitudes. This may be in part due to the very small number of match-ups at the latitude extremes. For the earlier years the confidence ranges are quite large and it is difficult to discern much trend in latitude. Figure 9 also shows which latitude bands have confidence intervals which do not overlap the intervals for the 2009 means when shifted to the same global mean as 2009 (highlighted in red). In the earlier years it is difficult to assess whether there is a zonal trend that is not so prominent in 2009. Through the AATSR period it is also unclear whether a trend in zonal mean bias is reducing towards later years. In some instances, bands not highlighted as being significantly different than the 2009 means are then lower or higher than the confidence interval of 2009 in the following year.

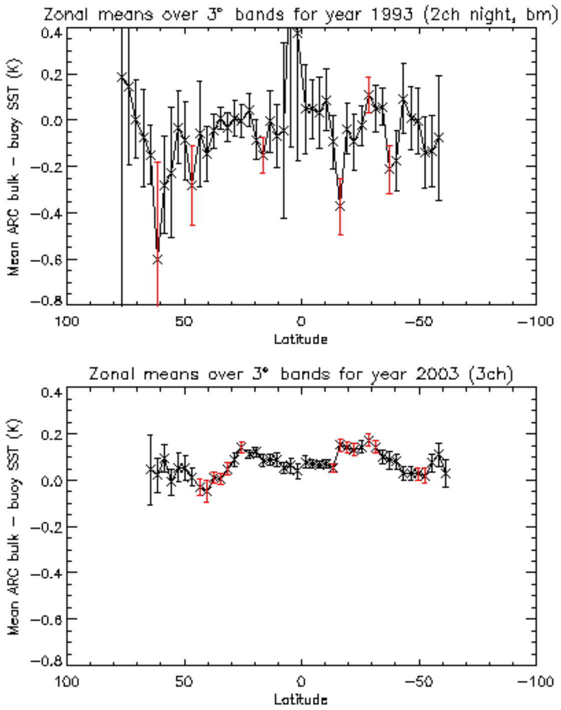


Figure 9 Selection of zonal means for 3 channel SSTs and 2 channel night-time (minimum Bayesian cloud mask) SST for 1993 (biases are averaged over 3° rings). Error bars show the 95% confidence interval. Those highlighted in red have confidence intervals that do not overlap the confidence intervals of the 2009 zonal means (calculated by shifting all points by the difference in the means between 2009 and the year being analysed).

4.1.5 Regional variation

From Figure 2 and from the discussion of zonal variation there is a strong suggestion that there are varying trends in the biases in the different ocean regions. The statistics of the different regions (listed in Table 1) were investigated and the year means are presented in Figure 10. This plot shows the results from the two channel night-time retrieval in order to include results from ATSR-1 but the three channel retrieval behaves in an almost identical way in the later years. The extreme value observed in the Tropical Atlantic in 1998 was caused by an unusually small number of collocations containing some very low values. A corresponding large peak in the standard deviation is also seen for this point and resulted in a very large confidence interval.

Generally, the spread in the mean biases and the standard deviations from the different regions decrease as the years increase. The transition to the ATSR-1 instrument is quite noticeable with a larger divergence of the means from the different regions. Throughout most of the time period, the North Atlantic and Southern Ocean regions tend to have lower mean biases – further evidence that SSTs around the higher latitudes often show cooler biases.

The two channel retrievals in daytime also show similar patterns (although the extreme value in the tropical Atlantic during 1998 is absent in the daytime). The daytime mean biases from the different regions do not converge as much in the later years and there is also a more distinct difference between lower mean biases in the higher latitude regions of the Southern Ocean and North Atlantic and the higher mean biases in the tropics. This could be partly due to different levels of insolation (discussed later in section 4.1.6).

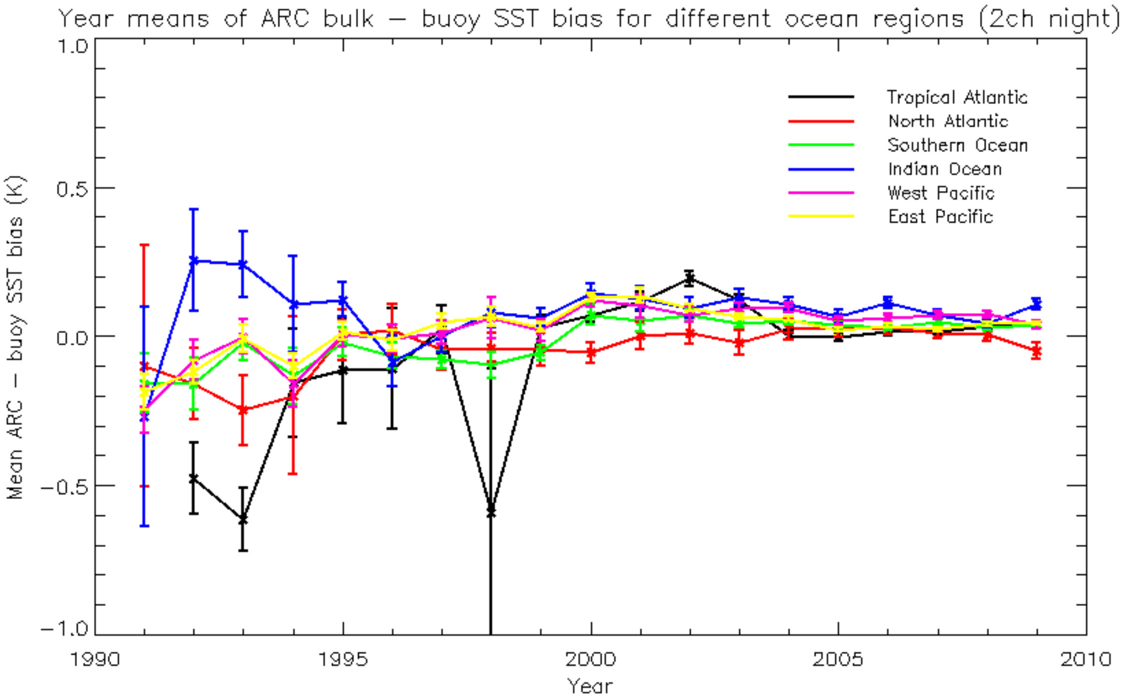


Figure 10 Plot comparing the year mean biases of different ocean regions using two channel night time SSTs and minimum Bayesian cloud flag for ATSR-1 period

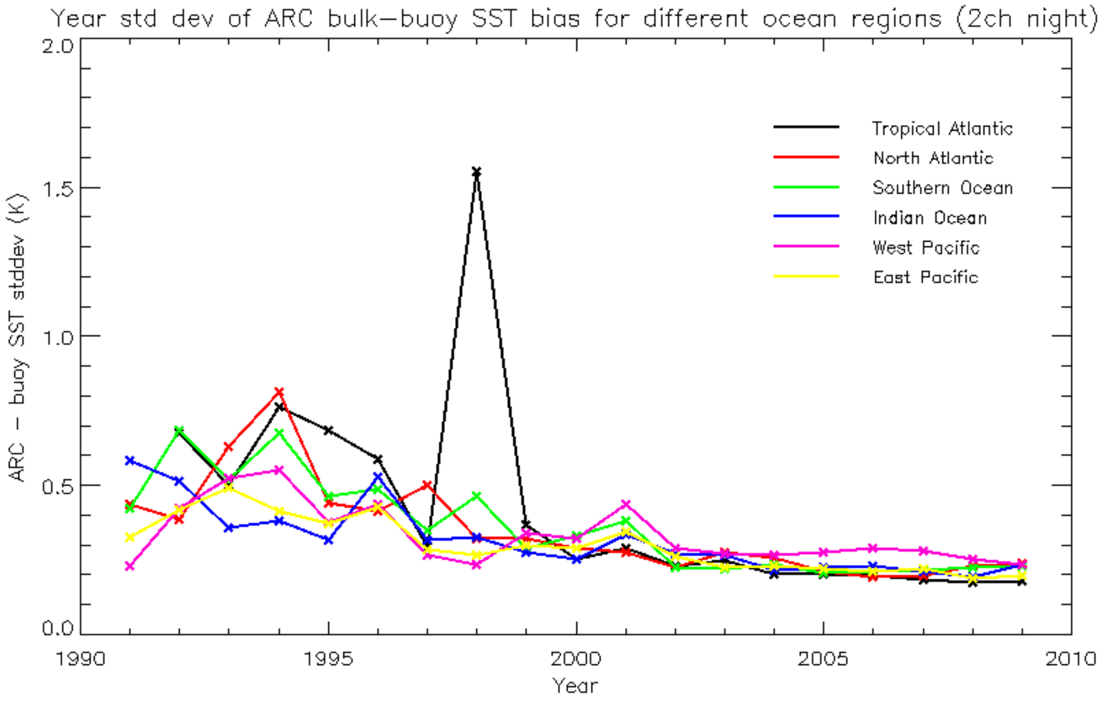


Figure 11 Plots of the standard deviation of the year biases for different ocean regions using two channel night-time SSTs and minimum Bayesian cloud flag for ATSR-1 period

4.1.6 Summary of dataset using Hovmoller plots

Figure 12 summarises some of the information discussed above regarding the difference in the data from different instruments and the zonal means for the three channel retrieval. The polar regions tend to have cooler biases for most of the time period while the mid-latitudes generally show warmer biases. The lack of data in the equatorial regions during 1997 and 1998 coupled with particularly cold biases in the north seem to be responsible for the overall low global means.

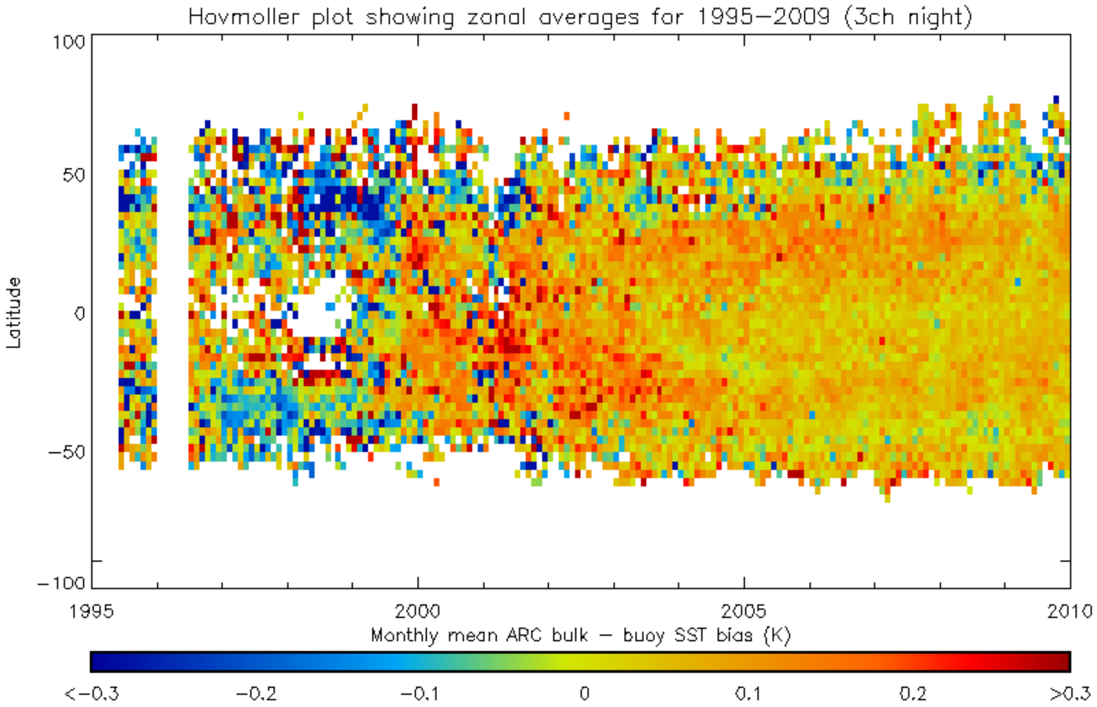
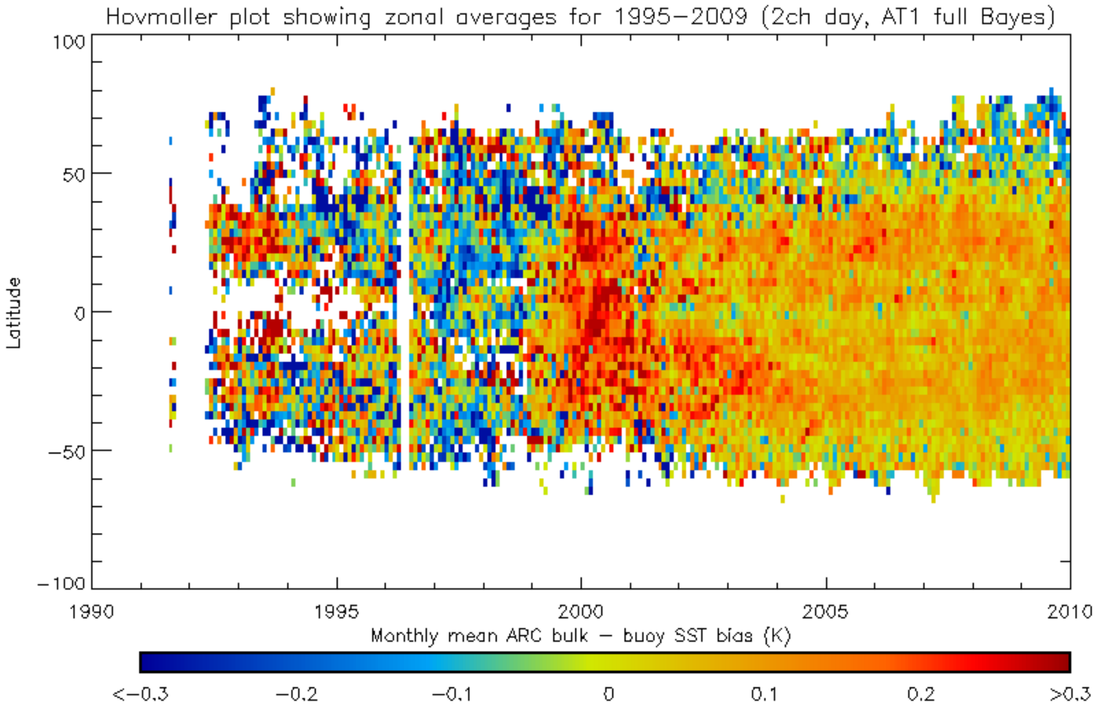


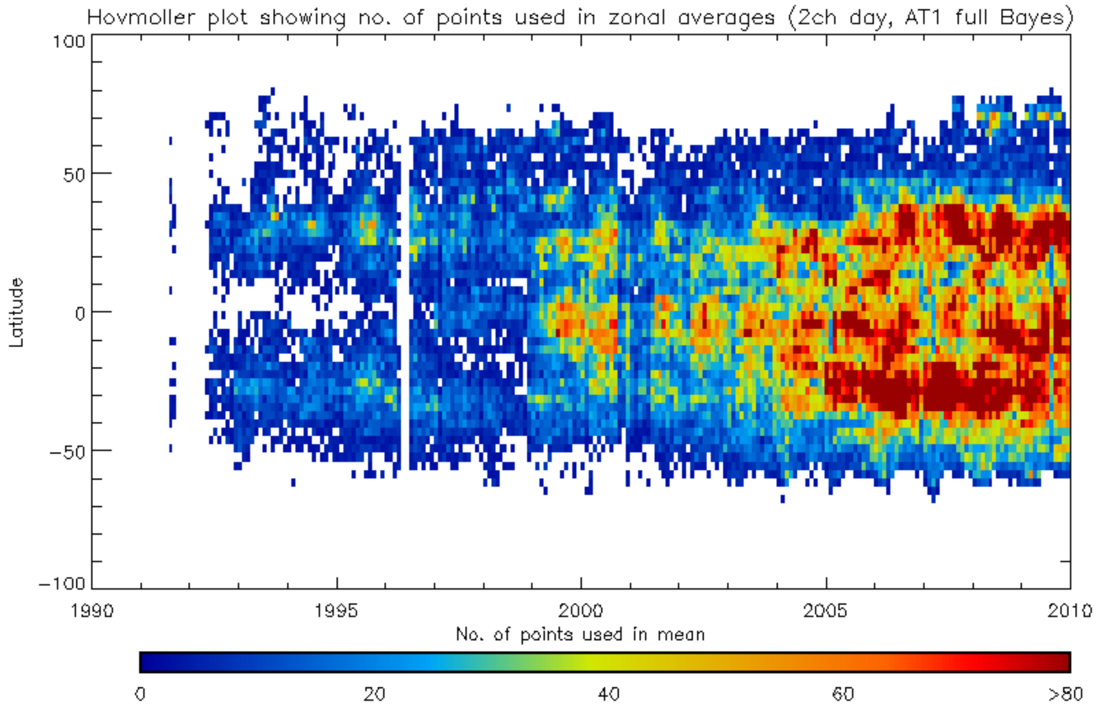
Figure 12 Hovmoller diagram showing the monthly zonal mean biases using averages over 3° bands of latitude

In order to demonstrate how the ATSR-1 period compares the two channel daytime retrievals are shown in Figure 13 (a) with the corresponding plot showing the number of collocations used for each mean in Figure 13 (b). The full Bayesian cloud mask has been used throughout. As with the three channel retrieval, the transition between the ATSR-2 and AATSR periods during 2002 is not clearly defined although there is still a noticeable difference in the stability to more extreme values in the earlier years. The global coverage is much lower in ATSR-1 than for the later years.

The number of collocations used for the calculation of each mean value in the Hovmoller plot shown in Figure 13 (b) is quite typical for the different retrieval types. It reveals clearly the decline in the number of collocations in the polar regions and the imbalance in numbers between the early and later parts of the dataset. The diagram also highlights the issue that statistics derived from data in the ATSR-1 period must be treated with more caution as the confidence intervals will be comparatively large.



(a)



(b)

Figure 13 (a) Hovmoller diagram showing the monthly zonal mean biases using averages over 3° bands. The full Bayesian cloud mask was used for all the data. **(b)** Hovmoller diagram showing the corresponding number of matchups used in calculating each mean

4.1.7 Relationships between SST bias and other variables (wind speed, insolation etc.)

Figure 14 shows some of the significant results from looking at the SST difference as a function of wind speed, insolation, total column water vapour (tcwv) and the uncertainty in the satellite measurement. As the ATSR-1 data were observed to be comparatively unstable it was decided to treat this part of the time series separately. The plots discussed below only used ATSR-2 and AATSR data but in each case the trend in the

ATSR-1 data was the same or the confidence intervals were too large to discern any dependence.

Figure 14 (a) and Figure 14 (b) show the dependence on the wind speed for the two channel daytime and three channel retrievals respectively. There is a slight trend for lower biases at higher wind speeds with a more noticeable trend for the daytime data - potential thermoclines should have been removed but this result could be associated with their occurrence at low wind speeds.

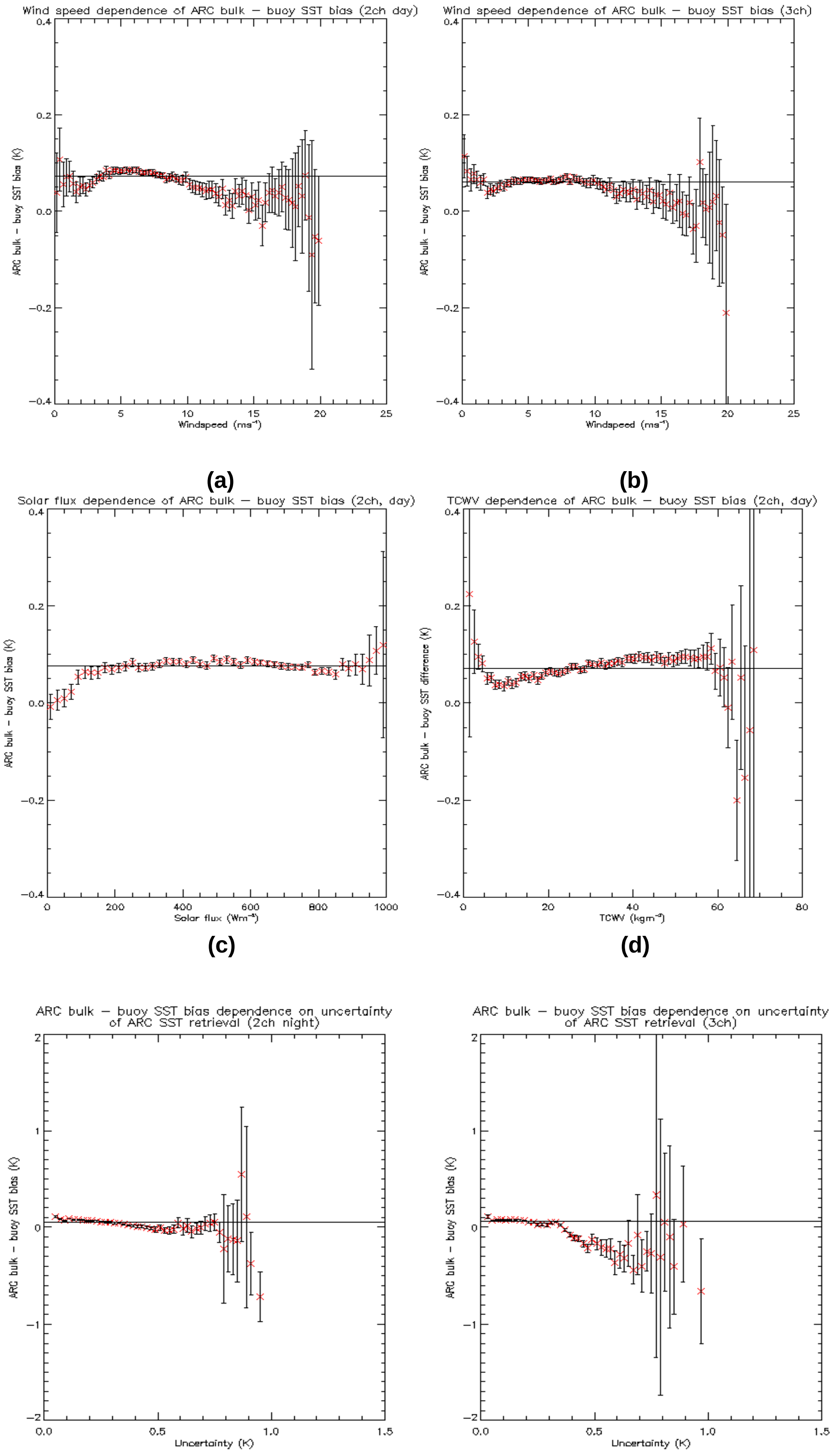
The wind data used in these plots are taken from the ERA-Interim reanalysis rather than measurements taken by the buoys. Analysis was attempted using the *in situ* observations but there were few measurements and an uneven spread of wind speeds making it difficult to make any conclusions. However, it is worth noting that the results did not contradict those seen with the model data.

The dependence of the two channel retrieval during daytime on the solar flux (data were provided by the ERA-Interim reanalysis) is quite small (Figure 14 (c)). A small decrease in bias can be seen for lower solar fluxes. The collocations in these bins are mostly found in the higher latitudes where lower biases have been observed (Figure 13 (a)). To assess whether there was any impact on the trend from any remaining potential thermoclines a stricter threshold of 0.05K for the difference between the sub-skin and bulk temperatures was used. However, there was no discernable difference seen in the results.

Figure 14 (d) shows a small trend of increasing ARC bulk – buoy SST bias as the total column water vapour (TCWV) increases for the two channel daytime retrieval. Dependences are not present for the night-time although for very high TCWV values there is a larger range of biases in a similar appearance to the variation seen in the extreme values in Figure 14 (d). The higher values of TCWV occur mainly in the tropics where the increased cloud cover can make retrievals more challenging. The larger amount of water vapour absorption may cause difficulty in calculating an accurate estimate of radiation from the sea surface. No relationship was found between the bias and the number of cloudy pixels in the field of view (at daytime or night-time).

The relationship between the bias and the uncertainty of the ARC SST retrieval was also considered. The two channel night-time (Figure 14 (e)) and daytime biases remain quite stable to an uncertainty value of around 0.6K and fluctuate for larger values. However, in the case of the three channel retrieval (Figure 14 (f)) the bias average only remains stable up to uncertainties of around 0.35K before decreasing for values up to around 0.6K and fluctuating beyond that. The locations of the match-ups with high values are distributed over the globe and in virtually all cases of uncertainties greater than around 0.32K the cells were adjacent to cloud edges (private communication, O. Embury, 2011). Removing the ARC SSTs with large uncertainty in the retrieval may improve the accuracy. However, the large majority of observations have retrieval uncertainties less than 0.4K so there is little impact on the overall statistics.

The dependence of the bias on the time difference between the buoy and satellite measurement was also investigated. There was no overall trend in the bias as the time increased.



(a) (b)

(c) (d)

(e) (f)

Figure 14 Dependences of the biases - the red points mark the mean for that bin and the black error bars show the 95% confidence interval for the mean of that bin. The overall mean of the biases used in the bins is also marked. No ATSR-1 data were included in these plots.

(a) Relationship between model wind speed and two channel retrieval in daytime, **(b)** Relationship between model wind speed and three channel retrieval at night-time **(c)** Relationship between model solar flux and two channel retrieval in daytime, **(d)** Relationship between total column water vapour (tcwv) and the two channel retrieval in daytime, **(e)** Relationship between uncertainty of ARC SST retrieval and the two channel night-time retrieval, **(f)** Relationship between uncertainty in ARC SST retrieval and the three channel retrieval

It is worth noting that throughout the analysis, although the bulk SST was used, statistics calculated using the ARC SSTs from 0.2m and 1.5m were not significantly different – the night-time retrievals using two or three channels showed very little change while the statistics for the daytime retrieval at 0.2m displayed the largest difference. Table 4 shows an example of the statistics using ARC SST corrected to different depths for the AATSR period data.

Channel selection	ARC – buoy SST bias ARC SST depth = 0.2m		ARC – buoy SST bias ARC SST depth = 1m		ARC – buoy SST bias ARC SST depth = 1.5m	
	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
AATSR 3ch night	0.059	0.143	0.059	0.143	0.059	0.143
AATSR 2ch night	0.053	0.159	0.053	0.159	0.053	0.159
AATSR 2ch day	0.079	0.158	0.071	0.157	0.054	0.163

Table 4 Mean and standard deviations of the global ARC – buoy SST bias using 1 grid boxes using data from the AATSR period corrected to different depths

4.2 Three-way error analysis

Five different experiments (defined in Table 2) were considered when carrying out the three-way error analysis – the first three vary the time difference between the ARC and buoy collocations from 60mins to 240mins while the last two experiments look at a larger distance over which the match-up can be considered successful. Two years were chosen for this more in-depth analysis – 2003 is the earliest complete year of AMSR-E data and also the year used in O'Carroll *et. al* (2008) so can be used for comparison while 2008 was also selected to find out how a later year with more data might perform. The three-way error analysis using the 180 minute window and same cell collocation was also carried out for the years 2003-2009 in order to assess whether there are any trends in the standard deviation of error for any of the instruments.

First considering 2003 and 2008 in the five different experiments listed earlier, Table 5 summarises the statistics for the different instruments. The mean and standard deviation for the ARC – buoy bias is lower in each experiment in the 2008 case but higher for the other instrument combinations. The buoy – AMSR-E biases are the lowest of the instrument combinations during 2003 but with the highest standard deviations in nearly all cases in both years. A large rise in the standard deviation is observed for the ARC – buoy bias in experiment 5 although the mean is reduced. The number of collocations also increases dramatically for the last experiment.

Table 6 shows the standard deviation of errors calculated for the different instruments. The same error is found for the ARC data in 2003 as was found for the AATSR data in the study of O'Carroll *et. al* (2008). Similar errors are also found for the AMSR-E SST in both studies while the error for the buoy SST in this report is slightly lower in comparison. The errors for the first three experiments, where only the time window varied, are reasonably consistent with similar values for both years. The ARC bulk SSTs produced the lowest standard deviation of error in each case. For the fifth experiment, although the error for the AMSR-E SST has not significantly changed, there is a large rise in the buoy and ARC errors. Allowing up to 1° separation between observations has led to a break down in the assumptions on which the three-way analysis is based. The error of representativeness (as discussed in O'Carroll *et. al* (2008)) must be negligible in order to gain reliable results but the inconsistency between the values in experiment 5 and the results of experiments 1 – 3 suggests this is no longer a reasonable assumption for the more relaxed spatial criterion. The numbers obtained in the final experiment cannot be taken as sensible values of the standard deviation of error. For experiment 4, where up to a 0.1° separation was permitted, the ARC bulk SST error has increased for both years indicating that the assumptions of the three-way analysis are starting to become less valid.

Expt. no.	Instrument combination	2003			2008		
		Mean (K)	Std. dev (K)	No. of matches	Mean (K)	Std. dev (K)	No. of matches
1	ARC – AMSR-E	0.081	0.488	6085	0.125	0.508	14808
	Buoy – AMSR-E	-0.001	0.505	6085	0.072	0.511	14808
	ARC - buoy	0.081	0.233	6085	0.052	0.202	14808
2	ARC – AMSR-E	0.088	0.504	4175	0.118	0.508	12666
	Buoy – AMSR-E	0.024	0.515	4175	0.065	0.512	12666
	ARC - buoy	0.064	0.222	4175	0.054	0.200	12666
3	ARC – AMSR-E	0.079	0.487	6482	0.124	0.509	15326
	Buoy – AMSR-E	-0.002	0.504	6482	0.072	0.512	15326
	ARC - buoy	0.080	0.235	6482	0.051	0.204	15326
4	ARC – AMSR-E	0.077	0.491	7755	0.123	0.510	18889
	Buoy – AMSR-E	-0.007	0.505	7755	0.067	0.509	18889
	ARC - buoy	0.084	0.252	7755	0.057	0.224	18889
5	ARC – AMSR-E	0.077	0.525	26165	0.111	0.560	65877
	Buoy – AMSR-E	0.027	0.686	26165	0.088	0.697	65877
	ARC - buoy	0.050	0.594	26165	0.024	0.610	65877

Table 5 Mean and standard deviations for the differences between ARC, buoy and AMSR-E SSTs using the 3 channel retrieval for the ARC bulk SSTs.

Expt. no.	Standard deviation of error for 2003 (K)			Standard deviation of error for 2008 (K)		
	ARC bulk SST	AMSR-E SST	Buoy SST	ARC SST	AMSR-E SST	Buoy SST
1	0.137	0.468	0.189	0.136	0.489	0.149
2	0.138	0.485	0.174	0.135	0.490	0.148
3	0.139	0.467	0.190	0.138	0.490	0.150
4	0.157	0.466	0.197	0.159	0.484	0.158
5	0.281	0.443	0.523	0.316	0.462	0.521

Table 6 Errors for the different observation types for the 3 channel retrieval

The three-way error analysis was also carried out for all AATSR years except 2002 using the criteria of experiment 1. The standard deviation of errors are shown in Table 7 and summarised in a graph in Figure 15. The values for the ARC SSTs have the smallest range over the 7 years and do not have any obvious trend. However, for the AMSR-E SSTs there seems to be a slight rise in error as the years progress which could be due to the instrument degrading with age. The buoy SSTs fall slightly in the early years then become more stable. The Data Buoy Cooperation Panel (DBCP) had a campaign to put out over 1250 drifting buoys to improve the network and this work was completed in 2005. The gradual introduction of these new buoys may be the cause of the decreasing error.

Instrument	Standard deviation of error for each year (K)						
	2003	2004	2005	2006	2007	2008	2009
ARC bulk SST	0.137	0.129	0.139	0.137	0.138	0.136	0.134
AMSR-E SST	0.468	0.462	0.462	0.466	0.482	0.489	0.500
Buoy SST	0.189	0.174	0.155	0.152	0.149	0.149	0.153

Table 7 Standard deviation of error for 2003 – 2009 for the ARC bulk, AMSR-E and buoy SSTs

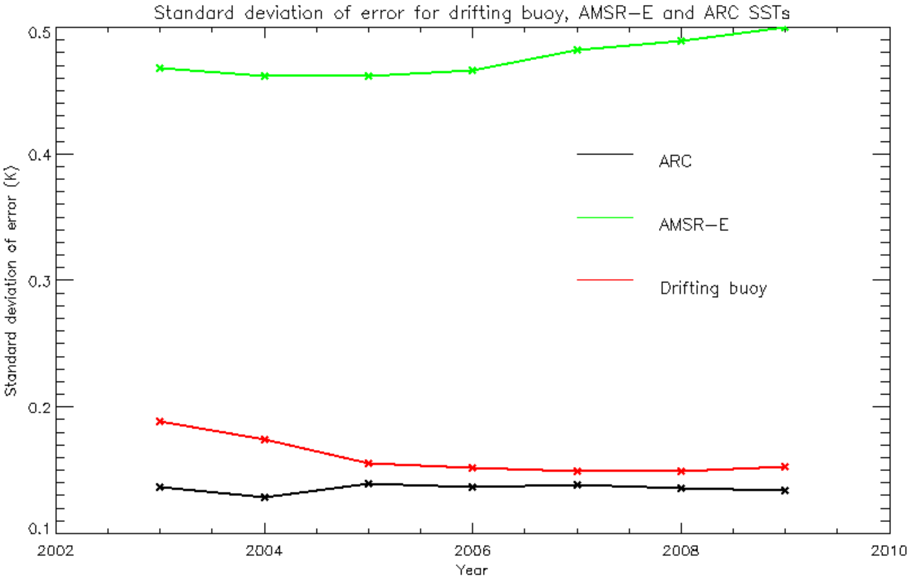


Figure 15 Plot of the standard deviation of error for 2003-2009 for the three instruments used in the three-way error analysis.

While analysing the match-ups using the three-way error analysis, data produced in the different experiments were also used to further investigate the sensitivity to the collocation criteria. Using data from 2002 – 2009, Table 8 shows the statistics for the different criteria. Changing the time window had only a small impact and the gain in accuracy of representing the satellite-buoy difference when using the smaller, 60 minute time window will most likely be lost due to the reduction in the number of collocations (by nearly 20% for 2002-2009). This particularly would not be desirable in the earlier years where there are already few match-ups available. Changing the spatial criterion to a threshold of 0.1° had relatively small effects on the mean bias while causing a small rise in the standard deviation. It is likely that the poorer resolution of the ICOADS did not have a large influence on the statistics. Increasing the allowed distance between the two observations further had the effect of lowering the mean bias. However, the more relaxed spatial criterion meant that the two instruments could potentially be measuring different locations with significantly dissimilar characteristics and this has led to a much high standard deviation. The global map of the biases showed more extreme values and greater regional variability. This highlights the fact that to gain more sensible results a small distance range should be used since this then selects the closest ARC estimate for the location of the buoy SST.

Expt. no.	No. of ARC – buoy match-ups	Mean global bias (3ch)	Standard deviation of global bias (3ch)
1	136649	0.061	0.143
2	112331	0.058	0.142
3	145487	0.060	0.142
4	178279	0.063	0.148
5	604445	0.020	0.311

Table 8 Summary of mean global biases and standard deviations for the 2002 – 2009 time period using different match-up criteria for the 3 channel retrieval

5. Conclusions

The ARC data have been tested using various statistical methods which included a three-way error analysis. The graphical results have shown a contrast between data using ATSR-1 and when using the other instruments. It was observed that the earlier years tend to contain much larger biases and show more variation. The ATSR-2 period showed a smoother transition between years but the AATSR era showed the greatest stability. The large increase in the number of collocations with time contributed greatly to reducing the confidence intervals on timescales from a week to a year. Overall, the SSTs using two channel night-time retrieval produced the lowest mean bias followed closely by the three channel retrieval and lastly the two channel daytime SSTs. However, the standard deviation of the bulk ARC – buoy SST difference was lowest for the three channel retrieval.

Regional variation was observed for the different retrievals – the lower latitude regions tended to show warmer biases. Higher mean biases in the tropics were also generally observed in the zonal means with more variation in the polar regions caused by small numbers of collocations. Seasonal variation was also found when considering the weekly means in the Northern and Southern hemisphere with lower biases found in the summer.

The dependence of the bias on other fields such as wind speed was also considered. A slight relationship between the wind speed and SST from day and night-time retrievals was found. The two channel SST bias was also found to increase with TCWV and the fall in bias at low levels of insolation corresponds to higher latitude locations where colder biases are more often observed. The relationship with the uncertainty on the satellite SST retrieval was also considered. It was also shown that for uncertainties greater than around 0.35K, the biases using the three channel retrieval could become less reliable while a higher threshold of 0.6K could be set for the two channel retrieval during both day and night.

The impact on the statistics due to the variation in buoy coverage throughout the time period was also important to consider. Until around 1997, the low density and poor global coverage meant that conclusions regarding the quality of the data were more difficult to make. Large confidence intervals, particularly for the ATSR-1 period and for the polar regions in some of the later years, indicated that the biases calculated were less reliable as a good representation of the true value. The impact on the statistics due to fewer buoys was also evident when considering smaller spatial or temporal resolutions such as the weekly mean biases or the zonal averages.

In the three-way error analysis, the ARC data for 2003 was shown to agree with the standard deviation of error that was found in the previous study of O'Carroll *et. al* (2008) when using the same criteria. Similar ARC errors were also found when considering subsequent years. The AMSR-E error appears to increase as the years progress, possibly due to instrument degradation, and the drifting buoy error decreases initially before reaching a more stable value after 2005. For experiments where the time window in the match-up was varied, errors for the three data sources were consistent across the two years, between experiments and close to values found in the previous study. Allowing a slightly greater distance between the buoy and ARC observations of 0.1° changed the mean and standard deviation by small amounts while the standard deviation of error increased as the assumption of negligible error of representativeness in the three-way error analysis became less valid. Generally, the statistics are reasonably insensitive to small changes in the match-up criteria.

However, the final experiment, in which the distance threshold was set at 1°, showed significantly higher standard deviation of errors for the ARC and buoy SSTs. The assumptions used by the three-way error analysis had now broken down leading to the values produced no longer being sensible. The error of representativeness provides a contribution to the overall error by accounting for the difference in the space/time scales of the measurement and in the analysis. The increase in the allowed spatial separation between the two observations caused the disparity in these scales to become significant enough that it was no longer reasonable to ignore the error arising from their difference. Relaxing the spatial criteria also had more impact on the means and standard deviations of the ARC – buoy bias than the change in time window. The characteristics of the ocean generally vary over relatively large time scales so continuing to increase the time over which the two observations can still be collocated should eventually lead to a similar breakdown in the three-way error analysis and greater differences in the mean bias and standard deviation of bias.

Overall, the three-way error analysis showed that the ARC SSTs from the AATSR era are of high accuracy but the graphical analysis revealed differences between the biases of the early and later years. However, in some cases it was difficult to judge the quality of the ATSR-1 data due to the lack of collocations resulting in large confidence intervals. Some regional and zonal biases also remained as well as small dependences on wind speed and TCWV.

Acknowledgements
AMSR-E data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science MEaSUREs DISCOVER Project and the AMSR-E Science Team. Data are available at www.remss.com.

Bibliography

1. Embury, O., https://www.wiki.ed.ac.uk/display/arcwiki/Test+Data, University of Edinburgh. Visited Sept 2011.
2. Knight, W., http://www.math.unb.ca/~knight/utility/t-table.htm, University of New Brunswick. Visited Feb 2010
3. Llewellyn-Jones, D.T., Corlett, G.K., Mutlow, C.T., 2007. AATSR - Completing the first 15 years global SST for climate, European Space Agency*, (Special Publication) ESA SP, (SP-636)*
4. Llewellyn-Jones, D.T., Feb 2006, AATSR Principal Investigator, Space ConneXions Contract: 2004-07-001/CPEG10
5. Merchant, C.J. et al., Deriving a sea surface temperature record suitable for climate research from the along track scanning radiometers, *J. Adv. Space Res. (2007), doi:10.1016/j.asr.2007.07.041*
6. 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, Vol. 25, No. 7, pp 1197-1207.*

Appendix A – Comparison of full Bayesian and SADIST cloud masks

A short assessment of the difference in ARC SST retrieved when using the full Bayesian or SADIST cloud mask was carried out using data from 2005 – 2009. The SADIST mask is currently used operationally in the Met Office ATSR product so it is useful to see how the improved cloud mask compares.

Figure 16 shows the difference in SST for the two channel night-time retrieval when using the different cloud masks. When using the Bayesian cloud mask, the SSTs appear slightly warmer in the tropics. This is supported by Figure 17 where the average of the 3° bands over the whole 5 year period is shown. This could be due to undetected cloud causing cooler SSTs when using the SADIST mask. A smaller magnitude effect is seen in the daytime two channel retrieval (not shown here). However, this effect is not present when looking at the three channel night-time SSTs (not shown here). This could be due to the three channel retrieval being more resilient to undetected cloud. The effects of residual cloud are greater in the longer wavelength channels of (A)ATSR so in the three channel retrieval, where information at a shorter wavelength is added, provided the retrieval coefficients are positive (this is an assumption and not necessarily true) then more weight will be given to this shorter wavelength in the SST retrieval leading to an overall smaller impact due to the cloud. However, a smaller magnitude difference in the tropics is not obvious in the plots of the three channel retrieval with the 95% confidence overlapping for most latitude bands.

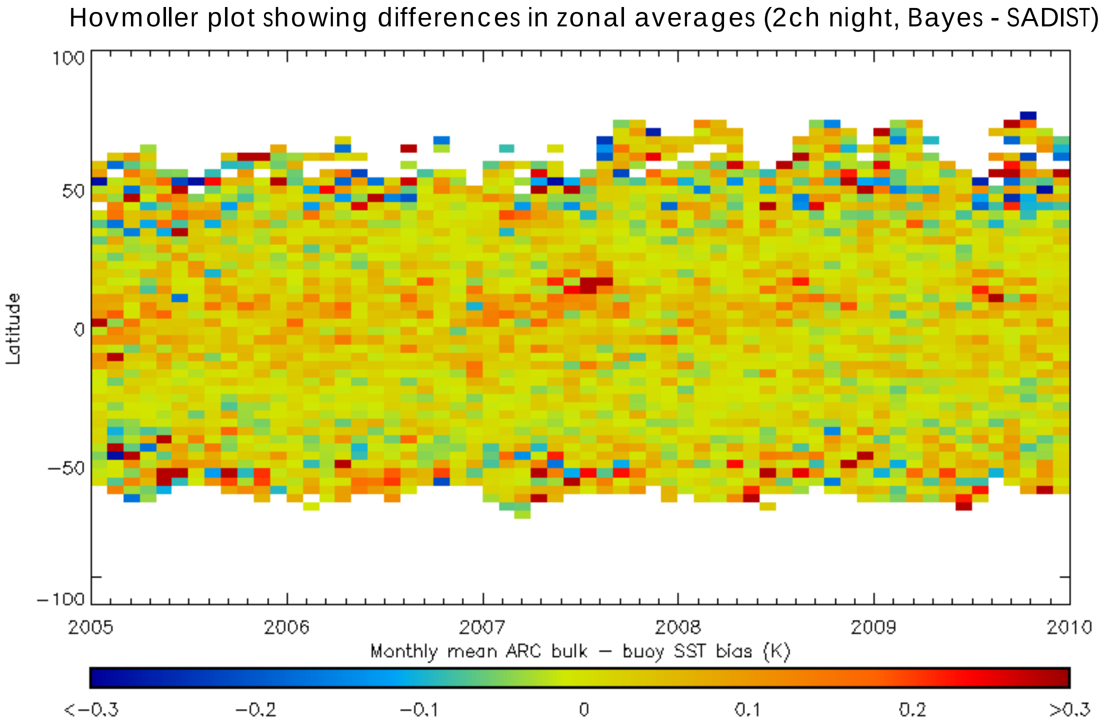


Figure 16 Hovmoller style plot showing the difference in the SST monthly zonal averages (using 3° bands) for the full Bayesian and SADIST cloud masks using two channel night-time retrievals

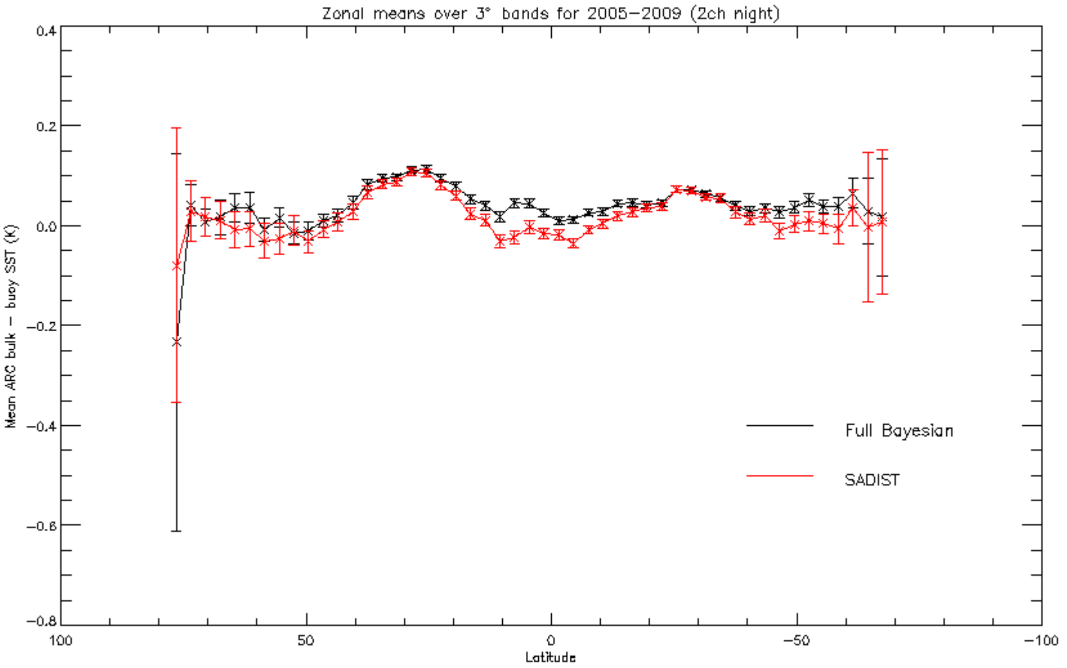


Figure 17 Plot showing the zonal mean ARC bulk – buoy difference using 3° bands with the same band averaged over 2005-2009 using two channel night-time SSTs. In red: SADIST cloud mask used, black: full Bayesian cloud mask used.

Met Office
FitzRoy Road, Exeter
Devon EX1 3PB
United Kingdom

Tel (UK): 0870 900 0100 (Int) : +44 1392 885680
Fax (UK): 0870 900 5050 (Int) :+44 1392 885681
enquiries@metoffice.gov.uk
www.metoffice.gov.uk