

Article

Quality Assessment of Sea Surface Temperature from ATSRs of the Climate Change Initiative (Phase 1)

Christoforos Tsamalis *  and Roger Saunders

Met Office, Fitzroy Road, Exeter EX1 3PB, UK; roger.saunders@metoffice.gov.uk

* Correspondence: christoforos.tsamalis@metoffice.gov.uk

Received: 30 January 2018; Accepted: 16 March 2018; Published: 21 March 2018



Abstract: Sea Surface Temperature (SST) observations from space have been made by the Along Track Scanning Radiometers (ATSRs) providing 20 years (August 1991–April 2012) of high quality data. As part of the ESA Climate Change Initiative (CCI) project, SSTs have been retrieved from the ATSRs. Here, the quality of CCI SST (Phase 1) from ATSRs is validated against drifting buoys. Only CCI ATSR SSTs (Version 1.1) are considered, to facilitate the comparison with the precursor dataset ATSR Reprocessing for Climate (ARC). The CCI retrievals compared with drifting buoys have a median difference slightly larger than 0.1 K. The median SST difference is larger in the tropics (~ 0.3 K) during the day, with the night time showing a spatially homogeneous pattern. ATSR-2 and AATSR show similar performance in terms of Robust Standard Deviation (RSD) being 0.2–0.3 K during night and about 0.1 K higher during day. On the other hand, ATSR-1 shows increasing RSD with time from 0.3 K to over 0.6 K. Triple collocation analysis has been applied for the first time on TMI/ATSR-2 observations and for daytime conditions when the wind speed is greater than 10 m/s. Both day and night results indicate that since 2004, the random uncertainty of drifting buoys and CCI AATSR is rather stable at about 0.22 K. Before 2004, drifting buoys have larger values (~ 0.3 K), while ATSR-2 shows slightly lower values (~ 0.2 K). The random uncertainty for AMSR-E is about 0.47 K, also rather stable with time, while as expected, the TMI has higher values of ~ 0.55 K. It is shown for the first time that the AMSR-E random uncertainty changes with latitude, being ~ 0.3 K in the tropics and about double this value at mid-latitudes. The SST uncertainties provided with the CCI data are slightly overestimated above 0.45 K and underestimated below 0.3 K during the day. The uncertainty model does not capture correctly the periods with instrument problems after the ATSR-1 3.7 μm channel failed and the gyro failure of ERS-2. During the night, the uncertainties are slightly underestimated. The CCI SSTs (Phase 1) do not yet match the quality of the ARC dataset when comparing to drifting buoys. The value of the ARC median bias is closer to zero than for CCI, while the RSD is about 0.05 K lower for ARC. ARC also shows a more homogeneous geographical distribution of median bias and RSD, although the differences between the two datasets are small. The observed discrepancies between CCI and ARC during the period of ATSR-1 are unexplained given that both datasets use the same retrieval method.

Keywords: ATSRs; sea surface temperature; CCI; ARC; validation; drifting buoys; AMSR-E; triple collocation

1. Introduction

Sea Surface Temperature (SST) is an Essential Climate Variable (ECV) for which there are available observations continuously since 1850 [1,2] made mainly by ships, but also from drifting and moored buoys during the recent decades. SST is directly related to and often dictates the exchanges of heat, momentum and gases between the ocean and the atmosphere [3], making it an important geophysical parameter for climate variability monitoring and prediction, operational weather and ocean forecasting,

ecosystem assessment and military operations [4]. Because of its significance, SST observations have been made operationally from space using the AVHRR instruments since 1981, as they offer the advantage of global coverage in contrast to in situ measurements [5,6]. Other satellite instruments designed for SST retrievals are the ATSRs, with ATSR-1 on board ERS-1 from August 1991 and its successor instruments ATSR-2, and AATSR, on board ERS-2 and ENVISAT, respectively, providing 20 years of high quality global SST observations [7]. Indeed, the three ATSRs SSTs outperform not only in comparison to other satellite instruments like AVHRRs, but also, their SST retrievals are of about equal or even better quality to in situ instruments [8,9]. Thus, the measurements from ATSRs have been used to assess the quality of in situ SST observations [10], to bias correct other satellite SST retrievals either directly [11,12] or through an SST analysis system [13], and to estimate the background error covariance parameters in SST analysis systems [14].

The ATSRs produce accurate SST retrievals thanks to their well-calibrated blackbodies and the low-noise infrared detectors cooled by a pair of Stirling-cycle coolers [15], while their dual view capability allows for better cloud detection and correction of atmospheric absorption and aerosols [16–18]. All three ATSRs observed Earth with four channels at 1.6, 3.7, 10.8 and 12 μm , while ATSR-2 and AATSR had three additional visible/NIR channels. The pre-launch calibration of the ATSRs demonstrated that the radiometric noise was below 0.05 K (at a reference of 270 K) in all thermal IR channels and was stable throughout the lifetimes of ATSR-2 and AATSR, and although variable for ATSR-1, the total drift of the mission was only 0.1 K [19]. The temporal stability of AATSR and its good radiometric calibration, especially for the 10.8 μm channel, have been verified against the high spectral resolution interferometer IASI on MetOp-A, which is the reference used by GSICS [20,21]. Similarly, ATSR-2's good radiometric calibration consistency to the level of 0.1 K has been verified by comparing with the high spectral resolution spectrometer AIRS [21].

The CCI SST [22], part of the ESA's Climate Change Initiative (CCI) [23], is an effort to produce a complete and homogeneous dataset of SST designed specifically with the climate quality criteria in mind, i.e., high accuracy and stability, while at the same time providing uncertainties per pixel. Phase 1 of CCI SST covers the period 1991–2012 when data from both ATSRs and AVHRRs are available. The project attempts to increase the global coverage on a daily scale by the use of AVHRRs, which have a larger swath width than ATSRs (2900 vs. 500 km), by combining their observations with the high quality SST retrievals based on ATSRs. The purpose of this study is to assess the quality of the SST retrievals from Phase 1 of the CCI project using comparisons with drifting buoys. The results will be compared with respective match-ups from the ATSR Reprocessing for Climate (ARC) dataset, which was the precursor of the CCI SST project [7,24,25]. As ARC involved only measurements from the ATSRs, in order to facilitate the comparison between CCI and ARC, hereafter, the focus is on the validation of CCI SST from ATSRs only. In Section 2, the datasets are described. Section 3 presents the assessment of CCI SST based on match-ups with drifting buoys, while in Section 4, similar results for ARC are given followed by the comparison with CCI SST. Finally, the conclusions are given in Section 5.

2. Datasets

2.1. Satellite SST Retrievals

The description of Phase 1 CCI SST project (CCI) is given by Merchant et al. [22] and the references therein, while for ARC, see Merchant et al. [7]. A short description is provided here focusing mainly on the differences and the similarities between CCI and ARC. In CCI, the retrieval of SST from ATSR-2 and AATSR brightness temperature observations is based on optimal estimation theory following the work of Merchant et al. [26]. In contrast, in ARC, the inversion of SST from ATSRs observations was based on retrieval coefficients estimated from radiative transfer simulations [27,28]. However, the SST from ATSR-1 in CCI also used the retrieval method of coefficients, in order to account for the impacts of the stratospheric aerosols from the Pinatubo eruption and the less stable performance of ATSR-1. Both CCI

and ARC SST retrievals are independent of in situ measurements in contrast to the majority of other satellite SST datasets, which are calibrated against them [29–32]. In both datasets, the cloud mask is based on a probabilistic (Bayesian) approach [33] (with minor differences concerning the versions of the radiative code and the NWP model), while dust aerosols are treated as clouds (i.e., masked) using the infrared dust index [34,35]. The ARC dataset provides more flexibility with SST retrievals calculated from both nadir and dual views either for two or three (during night time) channels, while CCI retrievals are based on dual view only, because of the better atmospheric correction [7,25], with two channels during daytime and three channels during night time. An exception to this is the period for ATSR-1 after the failure of 3.7 μm in May 1992, when the two 10.8 and 12 μm channels are used for both day and night retrievals. In addition to the skin temperature, which is the SST retrieved by the IR radiometers, both ARC and CCI provide the SST at a depth of 20 cm (ARC also provides depths at 1 and 1.5 m). The depth SST is based on a parameterization of the ocean skin effect [36] in order to facilitate the comparison with drifting buoys and the creation of climate datasets merging satellite and in situ observations [37]. In CCI, the depth SST is given at a standardized local time of 10:30 a.m./p.m. in contrast to the actual time of observation in ARC. The Equator crossing time for ATSR-1/2 was 10:30 a.m./p.m. and 30 min earlier for AATSR, stable (deviations less than 2 min) for all three satellite instruments during their lifetime [38]. Finally, the spatial resolution of the two datasets is different with ARC being 0.1° and CCI 0.05°, for the L3U product, which is the assessed product here (L3U stands for SST data from a single orbit file remapped and/or averaged onto a regular grid).

For the application of the triple collocation (or three-way error) analysis [39,40], in order to estimate the random uncertainty of CCI SSTs, use was made of SST retrievals from microwave (MW) observations obtained from TRMM Microwave Imager (TMI) [41] and Advanced Microwave Scanning Radiometer for EOS (AMSR-E) [42]. TMI was on board the TRMM satellite with a low-inclination orbit covering only the tropics and subtropics 40° S–40° N and providing good quality SST observations from 1998–2014 with variable Equator crossing times. AMSR-E on board the AQUA platform measured SST globally from 2002–2011 with an Equator crossing time of 1:30 am/pm. For both instruments, the SST retrieval developed by Remote Sensing Systems (Version 7) was used, with a spatial resolution of 0.25°. The SST retrieval algorithm for both TMI and AMSR-E is a physically-based two-stage regression that expresses SST in terms of the brightness temperatures for all the available channels of each instrument [42,43]. Here, TMI data are used from January 1998–July 2002 and AMSR-E data from August 2002–September 2011.

2.2. In Situ Observations and Quality Control

The assessment of the CCI dataset is based on comparisons with SST observations from drifting buoys available through International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Version 2.5 [44] for the period 1991–1996 and the Global Telecommunication System (GTS) for the period 1997–2012. Drifting buoys had a conservative quality control developed at the Met Office in order to eliminate buoys with gross errors [45]. The spatial collocation criterion between satellites and drifting buoys is 0.05° for CCI and 0.1° for ARC. In both cases, the maximum time window is 3 h, but when several observations from drifting buoys are available, the closest in time to the satellite overpass is chosen. It should be noted that the quality assessment of the ARC dataset provided by Lean and Saunders [25] was based mainly on the comparison of drifting buoy SST with ARC at a depth of 1 m. As the CCI depth is calculated only at 20 cm, the respective ARC depth (at 20 cm) will be used for consistency. However, according to the results of Lean and Saunders [25], the comparison statistics are similar for different SST depths, especially at night.

In both datasets, a common threshold has been applied to eliminate match-ups with a difference between buoys and ATSRs greater or equal to ± 3 K. In this way, problematic drifting buoy observations that have passed the above-mentioned quality control and/or cloud affected satellite SST retrievals were eliminated. Although, drifting buoys datasets exist with more elaborate quality control [44,46,47], here the same dataset as in Lean and Saunders [25] is used in order to provide a direct comparison

between CCI and ARC. The constant threshold has been chosen because the number of match-ups increases with time following the number of available drifting buoys. This can be seen in Figure 1a by the solid lines for day (red) and night (blue) when the number of match-ups before filtering increases from ~500 in 1991 to more than 30,000 in 2011. Note that the first data from ATSR-1 became available in August 1991, while AATSR ended in April 2012. The choice of the threshold is based on the annual standard deviation of the ATSRs-buoys unfiltered differences which is relatively close to 1 K for both datasets (Figure 1b). Thus, the selected threshold of 3 K is roughly a three sigma elimination of outliers, being mostly a conservative option that does not remove too many pairs.

The annual percentage of match-ups filtered out by the 3 K threshold is in general close to 1% for both datasets either for daytime or night time retrievals (Figure 1a, dashed lines). Nevertheless, there are years when the percentage of filtered match-ups is greater than 2%. For CCI these years are 1998–2001 and 2005, while for ARC (not shown) are 1998 and 2005. The years 1998 and 2005 appear in both datasets, with the percentage of collocations filtered out in 1998 during CCI daytime retrievals approaching 6%. The differences in the actual number of percentages between CCI and ARC indicate that the outliers are not only due to errors in drifting buoys. Furthermore, the difference in quality of drifting buoys from ICOADS and GTS can be clearly seen in Figure 1, with ICOADS buoys (1991–1996) having both smaller rejection percentage and standard deviation than the GTS data (1997–2012). In order to understand better the reason for the outliers, Figure 2a presents the bias (CCI minus drifting buoy SST) for daytime retrievals in 2005, the year with the larger number of outliers ($n = 568$). It can be observed that the outliers generally follow tracks rather than found in isolated locations or specific regions. This suggests drifting buoys with gross errors passed the conservative quality control. In an effort to locate possible regions that are prone to outliers, Figure 2b shows the geographical distribution of the daytime rejected match-ups for the whole period of CCI. For convenience, the number is provided only for 5° grid boxes with more than four rejected match-ups. Two regions that have systematic outliers (more than 30) for 2005 are the Mediterranean Sea and the Japan/East China Sea. In the next section, the location of the outliers will be reassessed in light of the CCI evaluation results.

Throughout this paper, the term bias is used, which by definition implies that drifting buoys provide unbiased observations. Certainly this is not true, at least not for all individual drifting buoys. Another study will focus on the quality of drifting buoys SST observations.

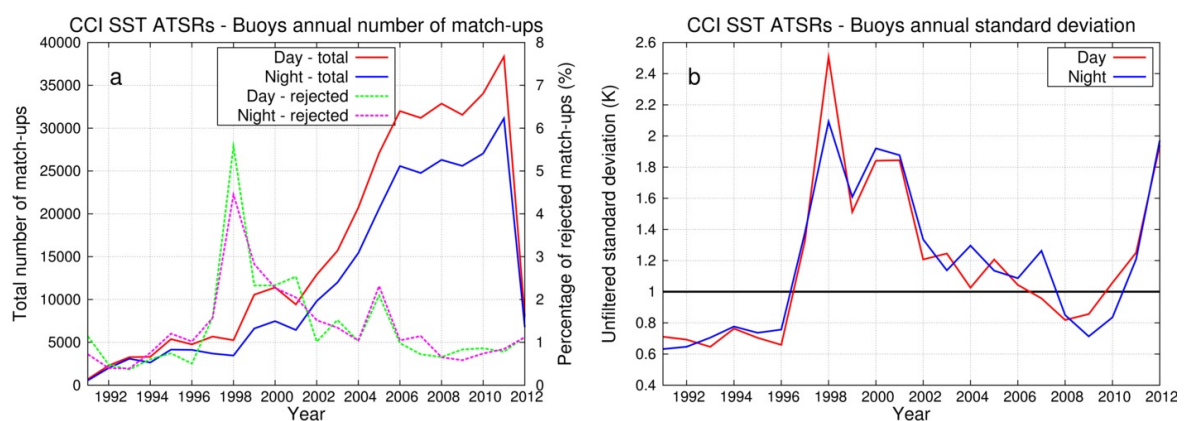


Figure 1. (a) Time series of the total number of match-ups (solid lines, left axis) before the application of ± 3 K filter during day (red) and night (blue) and of the percentage of eliminated match-ups due to the ± 3 K filter (dashed lines, right axis) during day (green) and night (magenta). (b) Annual time series of the standard deviation between CCI and drifting buoys before the application of ± 3 K filter during day (red) and night (blue).

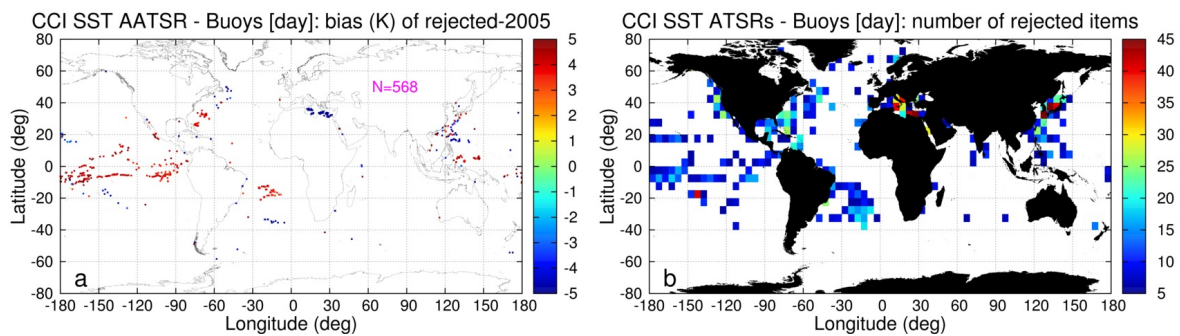


Figure 2. (a) SST difference of CCI minus drifting buoys for filtered match-ups in 2005 during daytime. (b) Number of CCI-buoys filtered match-ups for the period 1991–2012 in 5° boxes during daytime.

3. CCI SST Validation

3.1. Individual Performance

Firstly, the evaluation of each of the three ATSRs will be examined independently against drifting buoys. Table 1 presents the results for whole period with data for each instrument: ATSR-1 (2 August 1991–29 May 1996), ATSR-2 (1 June 1995–31 May 2003) and AATSR (24 July 2002–8 April 2012). Note that for ATSR-2, there are no data for the first six months of 1996 due to the scan mirror failure, while the quality of the data (when available) is reduced during the period mid-January to mid-June 2001 due to the gyro failure [7]. The mean and the median bias (CCI minus drifting buoys) is 0.10–0.13 K for all three instruments both for daytime and night time retrievals indicating a slight overestimation of depth SST by CCI. On the other hand, there are differences among the three instruments regarding the standard deviation or the Robust Standard Deviation (RSD) and between daytime and night time retrievals. During day, the RSD (standard deviation) is 0.55 (0.67) K for ATSR-1, 0.39 (0.53) K for ATSR-2 and 0.28 (0.43) K for AATSR. The respective values during night are 0.55 (0.69) K for ATSR-1, 0.28 (0.46) K for ATSR-2 and 0.22 (0.39) K for AATSR. The standard deviation is larger than RSD as it is more affected by pairs with bigger differences, while RSD better describes the distribution of the SST differences (RSD is 1.4826-times the median absolute deviation). Daytime retrievals have larger values than night time, unsurprising given the use of the 3.7 μm channel during night, which offers a better correction for atmospheric absorption, especially regarding water vapour in the Tropics. The 3.7 μm channel in ATSR-1 worked only for the first 10 months, meaning that the two channels retrieval was used for both day and night for the rest of its lifetime. This explains the absence of a difference at the standard deviation or RSD between day and night for ATSR-1. The performance of the instruments improves with time with ATSR-1 having the worst results and AATSR the best in terms of RSD and standard deviation. However, the apparent improvement with time from Table 1 can also reflect an improvement in terms of performance of the drifting buoys. The results of Table 1 are similar to those reported in Merchant et al. [22], although both the median bias and the RSD are slightly larger here. This probably reflects the less stringent quality control applied here for the drifting buoys, given that the number of match-ups is lower in the study of Merchant et al. [22].

In order to eliminate the variable performance of drifting buoys with time in the evaluation of ATSRs, Table 2 presents the statistics for the periods when the nominal operations of two instruments overlap. During the last seven months of 1995, both ATSR-1 and ATSR-2 were taking observations in tandem. The daytime results of the common period (Table 2) are similar to above-mentioned results for the whole lifetime of ATSR-1 and ATSR-2, with the exception of the mean bias for ATSR-2, which is reduced to 0.05 K. This confirms that ATSR-2 was performing better than ATSR-1. The common period between ATSR-2 and AATSR lasted for about 10 months. The daytime retrievals of ATSR-2 between August 2002 and May 2003 have a similar quality to the whole period (Table 1), but for AATSR both

the standard deviation and the RSD are increased approaching the values of ATSR-2. The opposite can be observed for the night time retrievals with AATSR showing the same performance to the lifetime results and the ATSR-2 standard deviation/RSD improving and matching the respective values of AATSR (Table 2). This fact indicates that ATSR-2 and AATSR CCI SSTs are essentially of similar quality, with AATSR being only very slightly better.

Table 1. Statistics (mean bias, standard deviation, median bias, Robust Standard Deviation (RSD) and match-ups number) of the comparisons between CCI ATSRs and drifting buoys. The period with available data for each instrument is indicated in parentheses, while daytime and night time retrievals are treated separately. [†] No data for ATSR-2 during the first 6 months of 1996.

Algorithm	Mean Bias (K)	St. Deviation (K)	Median Bias (K)	RSD (K)	Number
<i>ATSR-1 (08/1991–05/1996)</i>					
Day	0.13	0.67	0.13	0.55	16,241
Night	0.10	0.69	0.12	0.55	13,832
<i>ATSR-2 (06/1995–05/2003) [†]</i>					
Day	0.10	0.53	0.10	0.39	66,029
Night	0.11	0.46	0.12	0.28	45,861
<i>AATSR (07/2002–04/2012)</i>					
Day	0.11	0.43	0.12	0.28	274,795
Night	0.10	0.39	0.11	0.22	217,622

Table 2. As in Table 1 but for the common periods (indicated in parentheses) when observations from two ATSRs are available. For ATSR-1/2, only daytime comparisons are shown, because the night time retrieval algorithms are different due to the early failure of ATSR-1 3.7 μm channel.

Algorithm	Mean Bias (K)	St. Deviation (K)	Median Bias (K)	RSD (K)	Number
<i>ATSR-1 (06/1995–12/1995)</i>					
Day	0.11	0.71	0.12	0.56	3609
<i>ATSR-2 (06/1995–12/1995)</i>					
Day	0.05	0.57	0.07	0.39	3791
<i>ATSR-2 (08/2002–05/2003)</i>					
Day	0.09	0.52	0.11	0.39	10,794
Night	0.14	0.40	0.15	0.25	8203
<i>AATSR (08/2002–05/2003)</i>					
Day	0.08	0.50	0.10	0.34	11,429
Night	0.12	0.40	0.12	0.25	8719

3.2. Geographical Distribution

Figure 3 presents the maps of median bias, RSD and number of collocations for the 20 years period separately for day and night. The median bias and RSD are calculated for every 5° grid box with at least 20 match-ups inside it for the whole period of ATSRs. Thus, the white boxes indicate lack of sufficient number of collocations. Regarding the overlapping periods between satellites (Table 2), data from the best performing satellite have been used i.e., ATSR-2 and AATSR (instead of ATSR-1 and ATSR-2). It can be seen that during day the CCI depth SST is significantly warmer than the buoys SST in the tropics ~ 0.3 K, while in general for the other regions, the median difference is inside the range $[-0.1, 0.1]$ K with CCI being mainly slightly warmer than the buoys. Median differences less than -0.1 K can be observed in the Black Sea and off the coast of Arabian Peninsula. On the other hand, during night time, the CCI is found again warmer than buoys around 0.1 – 0.2 K, but now, the difference

is more homogeneous spatially. Higher differences are found over the northern Indian Ocean and the Western Pacific. Especially, in the case of the north Indian Ocean, there is a sharp transition with the bias being cold off the Arabian Peninsula becoming warm close to the Indian Peninsula. A less significant gradient can be also observed over the eastern part of tropical Atlantic. The cold bias off western Africa and Arabian Peninsula may be related to dust aerosols, which are abundant in these two areas [48,49]. The location of the daytime warm bias in the tropics seems to be related to the absorption of IR radiation by water vapour, which is retrieved less accurately with the two-channel algorithm, used during the day. A very similar geographical pattern for the median difference between the CCI SST analysis and drifting buoys has been found in Merchant et al. [22].

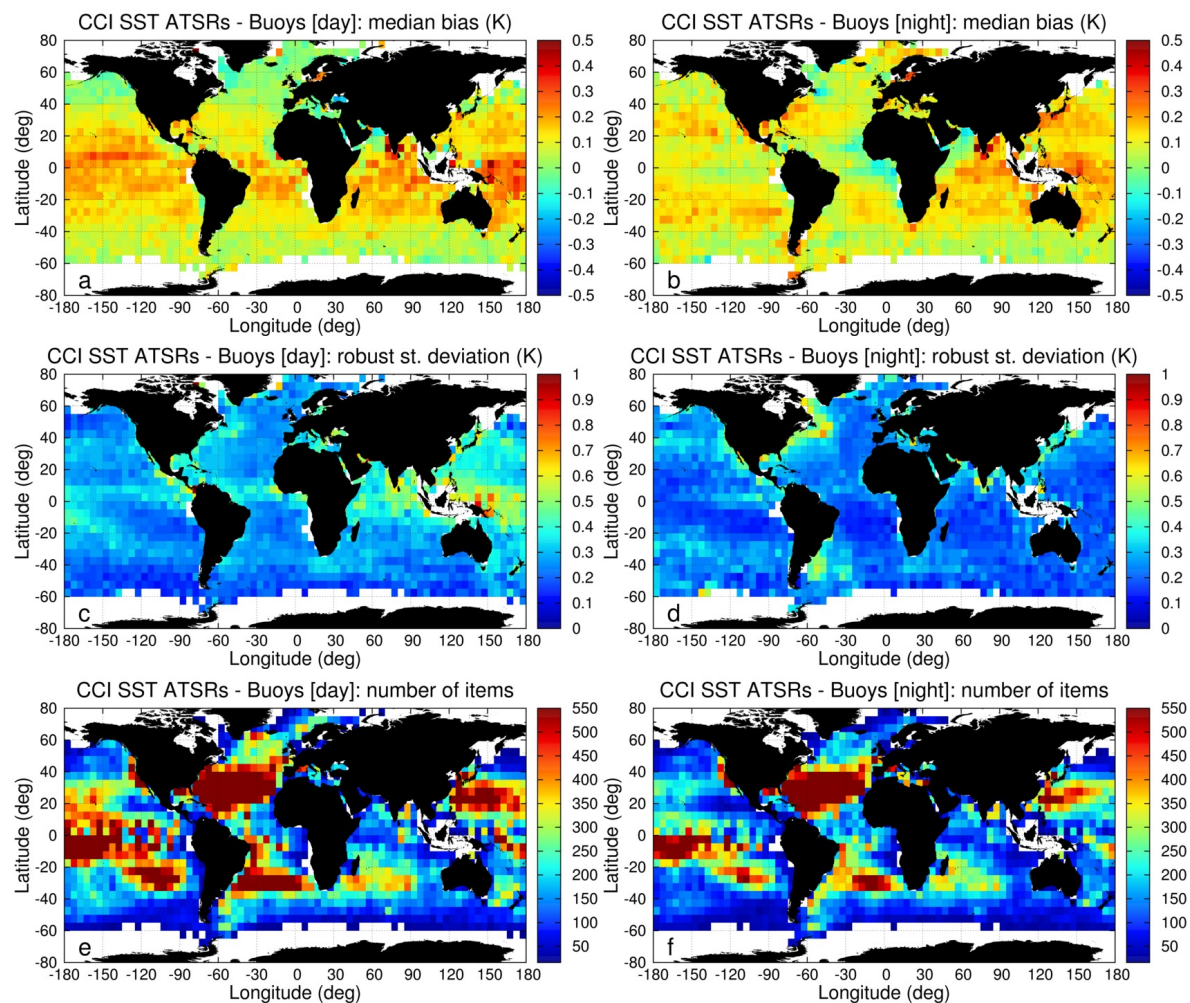


Figure 3. The median bias (a,b), robust standard deviation (c,d) and number of match-ups (e,f) of ATSRs' CCI SST against drifting buoys averaged in $5^\circ \times 5^\circ$ boxes for the period: August 1991–April 2012. The results for daytime (night time) observations are presented in left (right) column. Gridboxes with less than 20 match-ups appear white.

Regarding the robust standard deviation of the difference between CCI and buoys, during the day, larger values ~ 0.6 K are found over the Maritime Continent (SE Asia) and Northwest Pacific (Figure 3c). During the night, the spatial distribution is again more homogeneous than daytime and only off the eastern coast of Canada, and the south-western Atlantic Ocean RSD reaches 0.5 K (Figure 3d). The geographical distribution of match-ups is not spatially homogeneous (Figure 3e,f), but this fact does not impact the median difference or the robust standard deviation.

It should be kept in mind that the number of drifting buoys increased considerably with time (Figure 1a and Table 1), especially after 2003, meaning that the maps reflect mostly the performance of

AATSR. The same analysis has been repeated for each one of the three ATSRs (not shown); the results are similar to Figure 3a,c for ATSR-2, and the geographical patterns are very similar, although the values for the bias and the RSD are slightly larger. For ATSR-1, it is difficult to assess as there are far fewer collocations to give statistically-significant results about the geographical patterns of median bias and the RSD.

It can be useful to investigate if there is a specific geographical distribution for the collocations with the largest differences between CCI and drifting buoys. Figure 4 presents the mean bias and number of match-ups for the collocations with absolute difference larger than 1 K. The threshold of 1 K is chosen arbitrarily taking into account that collocations with difference greater than ± 3 K are discarded (Section 2), while an important number of match-ups should be left in order to arrive to some conclusions. The comparison of Figure 4a with Figure 3a indicates that the pairs with $|\Delta SST| > 1$ K during daytime contribute to the warm bias observed in the tropics and the western Pacific. For the other regions, the large differences from these collocations cancel out more or less giving a mean value close to 0 K. It is important to note that the number of the collocations per grid box with large SST differences is about 5–15 (Figure 4b) and much lower, at least an order of magnitude, than the total number of collocations (Figure 3e). For the night time retrieval (not shown), there are even fewer pairs with absolute SST differences larger than 1 K. The only geographic patterns emerging during night are the warm biases over the north-western Pacific and around the Indian Peninsula (similarly to Figure 3b). The identified regions with systematic large positive SST differences probably indicate deficiencies in the SST retrieval rather than issues with observations from drifting buoys.

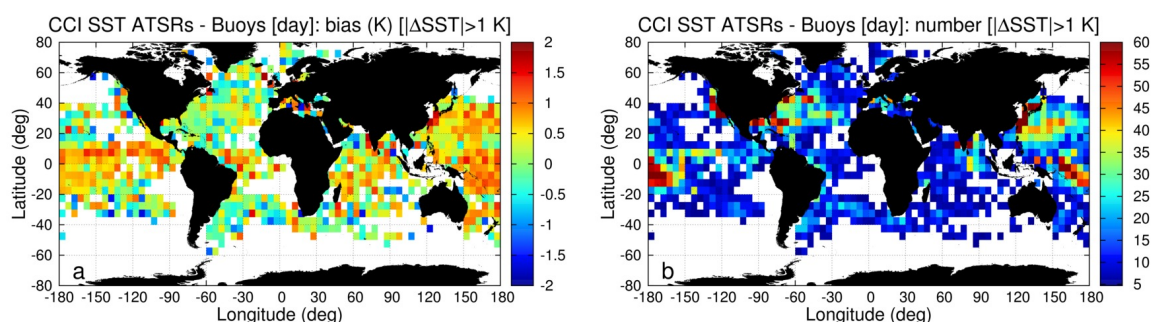


Figure 4. Maps of mean bias (a) and number of match-ups (b) for the collocations with $|\Delta SST| > 1$ K during daytime. Gridboxes with less than 5 match-ups appear white. Note the different scales to Figure 3.

3.3. Time Series

In order to have a clear indication of how the statistics are evolving with time given that Figure 3 reflects mostly the results of AATSR, Figure 5 presents the time series of CCI for the median bias and the RSD. The median bias is almost always positive with the exception of the last months of 1991 and the years 1997 and 1998 for the daytime retrieval only, when it reaches -0.1 K. From 2002 onwards, the daytime and night time retrievals have the same monthly global median bias of about 0.12 K, a fact clearly reflecting the stability of AATSR. Similar values of the median bias for the daytime and the night time retrievals are also found for the ATSR-1 period, but for ATSR-2 from 1995–2001, there are differences between day and night, although these are in general smaller than 0.1 K. The similar values between day and night for the ATSR-1 after 27 May 1992 (when the $3.7 \mu\text{m}$ channel failure occurred) are not surprising as the retrieval algorithms are the same, although the daytime cloud mask also uses the $1.6 \mu\text{m}$ channel. When switching between ATSR-1 and ATSR-2 (the ATSR-2 scan mirror failed in December 1995 and restarted in July 1996), an abrupt change of the median bias can be observed during night (for daytime observations a change also exists, but it is less significant). This is as expected as ATSR-2 uses the $3.7 \mu\text{m}$ channel in the night time algorithm. The median

bias of ATSR-1 is not stable temporally increasing from 1991–1993 reaching 0.3 K and then mainly decreasing towards 0 K. For ATSR-2, the changes are more of stepwise nature with the median bias being mostly stable between them, especially for the night time retrieval. These step changes occur around January 1997 (decrease from 0.15 K–0.05 K during night) and then around October 1999 (increase from 0.05 K–0.2 K during night), with the changes being slightly larger in magnitude for the daytime retrieval. The median bias for the whole 20 years of ATSRs CCI is almost always within ± 0.2 K with the exception of the periods 1993–1994 and 2000.

The temporal evolution of the robust standard deviation is simpler than the median bias (Figure 5b). Always, the night time RSD is lower than the daytime by about 0.1 K (for ATSR-1 after May 1992, there are no differences as the day and night retrieval algorithms are the same, using only the 10.8 and 12 μm channels). For ATSR-1, the RSD increases with time from 0.3–0.4 K to about 0.6–0.7 K with the most rapid change occurring during 1992. On the other hand, for ATSR-2 and AATSR, the RSD decreases slightly with time being lower for AATSR and the lowest values of ~ 0.21 K (0.28 K) for night time (daytime) obtained during 2008–2010. The switches from ATSR-1 to ATSR-2 are obvious with the RSD reducing from ~ 0.6 K to 0.3–0.4 K. However, it is important to note that the RSD of the first months of ATSR-1 (before the failure of 3.7 μm channel) is very similar to the values found during the first months of ATSR-2 operations. The switch from ATSR-2 to AATSR is smooth, especially for the night time retrieval. There is an apparent jump for the daytime RSD from 0.35 K to 0.6 K during February–May 2001 due to the gyro failure on ERS-2.

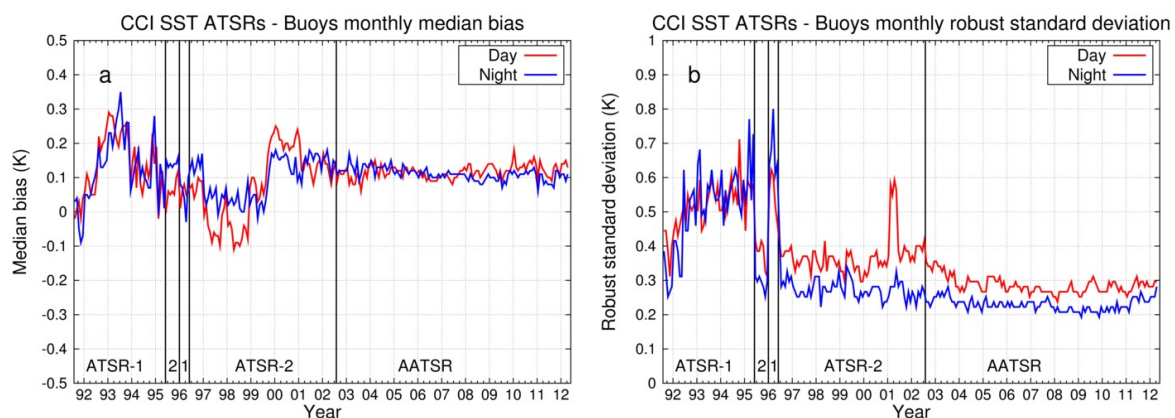


Figure 5. Global monthly time series of median bias (a) and robust standard deviation (b) of CCI against drifting buoys. The red (blue) line is for daytime (night time) retrievals. The black vertical lines indicate the switch of the time series from one instrument to another as indicated at the bottom of each panel.

Another way to investigate the temporal performance of the CCI against drifting buoys is by presenting how the percentages of match-ups for fixed SST threshold differences evolve. Figure 6 shows the evolution for differences less than 0.1 K or 0.2 K and larger than 1 K or 2 K. The thresholds are ad hoc, but the lowest values represent mostly the optimal cases while the largest thresholds are indicative for possible significant (i.e. in terms of magnitude) events. Similarly with the RSD (Figure 5b), there is a clear separation between night and day (again except for the period after the failure of 3.7 μm channel on ATSR-1), with the night retrievals showing better performance. This stratification between day and night retrievals breaks down for the $|\Delta\text{SST}| > 2$ K and the AATSR period for the differences larger than 1 K. Likewise for RSD, the performance improves with time as the percentage within 0.1 or 0.2 K increases, and the percentage larger than 1 K decreases. About 23% (30%) of day (night) retrievals are within 0.1 K from drifting buoys SST. The percentage of collocations within 0.2 K is almost always larger than 40% (50%) for day (night) retrievals. However, the percentages for the ATSR-1 period after the 3.7 μm channel failure are lower being around 14% (27%) within 0.1 K (0.2 K). Poor performance for the same period of ATSR-1 can also be observed in the percentage of match-ups

which are larger than 1 K (Figure 6b), reaching values larger than 20% for some months towards the end of the mission. Despite this ATSR-1 period, the 1 K percentage is around 5%. Interestingly, the percentage of collocations that are larger than 2 K does not fluctuate a lot with time being 0.5–1%, although the ATSR-1 period shows more variability from one month to the next. Regarding the period of the ERS-2 gyro failure (beginning of 2001), there is an obvious decrease of the percentages within 0.1 K or 0.2 K and an increase of the percentage larger than 1 K for the daytime retrievals. Nevertheless, the night time retrievals do not seem to be impacted significantly by the gyro failure. In addition, the gyro failure does not affect the percentage of collocations larger than 2 K, which means that any degradation of the CCI due to the gyro failure is within 2 K. In Figure 6b, a spike can be observed for both 1 K and 2 K percentages in August/September 2009, which does not coincide with any known instrumental issue for AATSR, thus indicating an issue with the drifting buoys or at least their quality control.

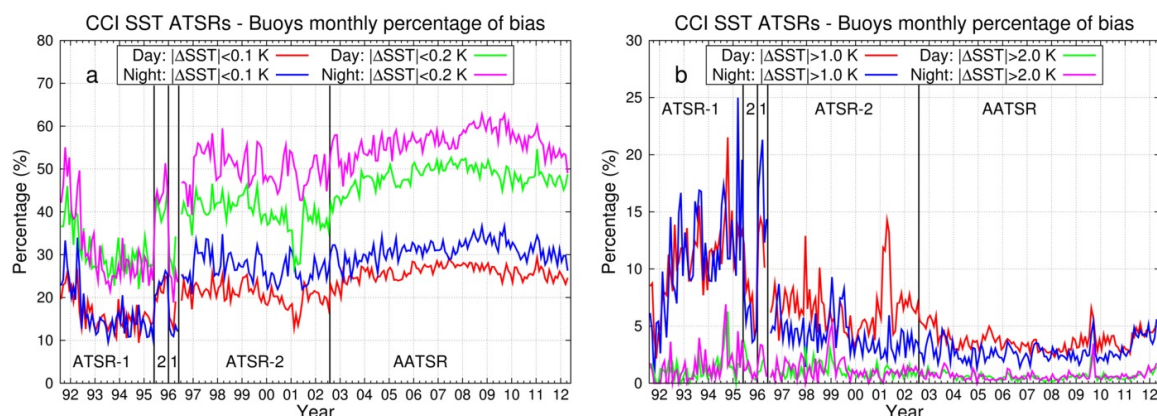


Figure 6. Temporal evolution of monthly percentages of CCI and drifting buoys SST absolute differences for various thresholds separately for daytime and night time retrievals. (a) Within 0.1 K and 0.2 K. (b) Larger than 1 K and 2 K.

3.4. Triple Collocation Analysis

The triple collocation or three way error analysis uses three independent datasets observing the same quantity (here SST) in order to calculate the standard deviation of the error for every one given the fact that the observations are uncorrelated [39,40]. If the assumptions behind the triple collocation hold, the standard deviation of the error is the random error for observations of each instrument (or more correctly the random uncertainty). Here, the latest version of AMSR-E SST is used (Version 7, which differs from the versions used by Lean and Saunders [25] and O’Carroll et al. [39]). For the first time to the best of our knowledge, the results are extended backwards in time by using the SST observations of TMI (Version 7.1) before August 2002. Thus, ATSR-2/TMI are used for the years 1998–2001 and AATSR/AMSR-E for the years 2003–2011, while for 2002, both sets are used. The match-ups of ATSRs with buoys are collocated with the closest observation of MW SST (provided with resolution of 0.25°) in a time window of 180 min for all three observations. As previously (Section 2), only collocations with absolute differences less than 3 K are used for any combination of the three datasets. In the past, the triple collocation approach has been used only for night time SST observations in order to avoid the impact of the diurnal thermocline [50–52]. However, here, for the first time to the best of our knowledge, the triple collocation is applied also during day, only when the wind speed observed from the MW imagers is larger than 10 m/s. This wind threshold is based on the results of previous studies [50,51,53,54], as it is expected that when the wind speed is above 10 m/s, the diurnal SST cycle should be absent especially for the ATSR overpass time of 10:30 LT.

Figure 7 presents the annual results of the triple collocation analysis. It can be noted that after 2003 both AATSR and drifting buoys have a stable standard deviation of the error ~ 0.22 K both during

day and night. For the period before 2004, the daytime results indicate that ATSR-2 and AATSR are at the same level, while the night time values suggest that ATSR-2 is slightly better. The drifting buoys before 2004 show a worse quality with values for the standard deviation of the error of 0.3 K or above for the daytime results. Regarding AMSR-E, the night time results indicate a rather stable value of ~ 0.47 K, but for the daytime conditions this value increases to about 0.6 K. The TMI has higher values than AMSR-E for the standard deviation of the error being 0.55 K during night and even larger during day. The poorer performance of TMI in comparison to AMSR-E is not surprising given that TMI lacks the 7 GHz channel, which is more sensitive to SST [42,43]. The daytime results before 2004 show a lot of annual variability for all three types of instruments. This variability coincides with a number of collocations less than 300 (cyan line) and even less than 100 for 1999 to 2001. Note that for 1998, there are not even 20 collocations during daytime to produce statistically meaningful results. The variability is expected as the number of collocations reduces to lower than about 500 [40], so the daytime results before 2004 should be interpreted with caution.

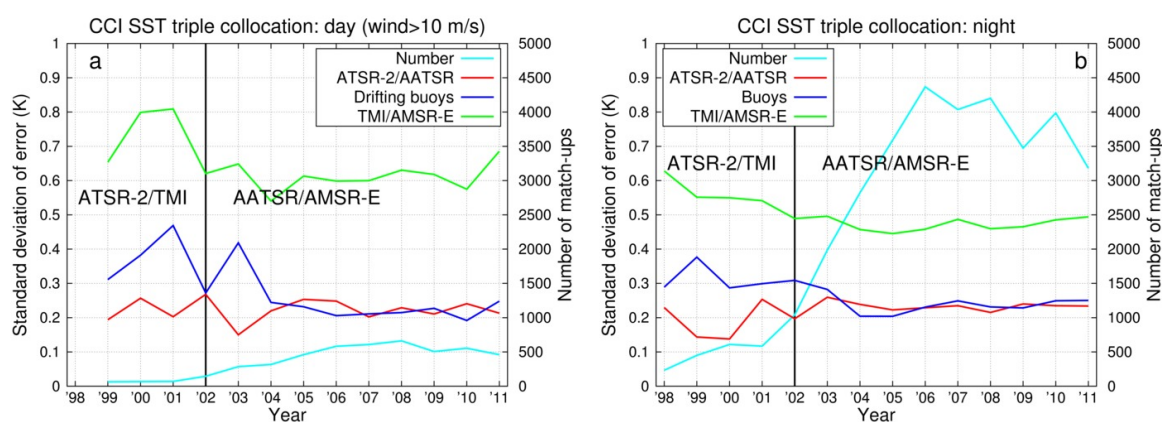


Figure 7. Triple collocation annual results. For the period January 1998–July 2002 observations from ATSR-2/TMI are used, while from August 2002–September 2011, AATSR/AMSR-E are used. The cyan line indicates the number of collocations (right axis), while the red, blue and green lines the standard deviation of the error (left axis) for ATSR-2/AATSR, drifting buoys and TMI/AMSR-E, respectively. (a) During daytime when the TMI/AMSR-E wind speed is larger than 10 m/s and (b) during night time.

It is important to know if the results of the triple collocation vary geographically. Figure 8 presents the Hovmöller diagrams for the standard deviation of the error of the three instrument types (IR imagers, in situ and MW imagers) together with the number of collocations. It can be seen that for the ATSRs (Figure 8a) and the drifting buoys (Figure 8b), the results are rather stable both in time and latitude with the majority of the grid boxes having values between 0.1 K and 0.3 K. The observed variability around the value of 0.22 K and mainly few values outside the range of [0.1, 0.3] K are the result of relatively low number of collocations, especially when the number is less than ~ 100 (Figure 8d). On the other hand, AMSR-E does show a variable behaviour with latitude with observations north of 35°S having a standard deviation of error ~ 0.3 K, while south of 35°S , the value doubles to 0.5–0.6 K. This characteristic performance of AMSR-E is stable with time. To the best of our knowledge, this is the first time that the latitudinal variability of AMSR-E SST is reported. While such a latitudinal dependence is expected for an MW SST retrieval using the 11-GHz channel [43], this should not be the case for the 7-GHz channel, which is available on AMSR-E. Regarding TMI, the standard deviation of error is more stable with values 0.5–0.6 K, similar to what is found in Figure 7b. There is also an indication that TMI performance degrades south of 35°S , but as the number of collocations is small (about 60) for this latitude, it is hard to arrive at firm conclusions. The small number of collocations can sometimes prevent the application of the triple collocation approach, e.g., the white grid boxes in Figure 8 for 2010 and 2011 in northern tropics. It is important to note that the triple collocation during night samples mainly the Southern Hemisphere mid-latitudes (Figure 8d). This fact

and the variable performance of AMSR-E with latitude could explain the differences seen in Figure 7 between day and night for this instrument.

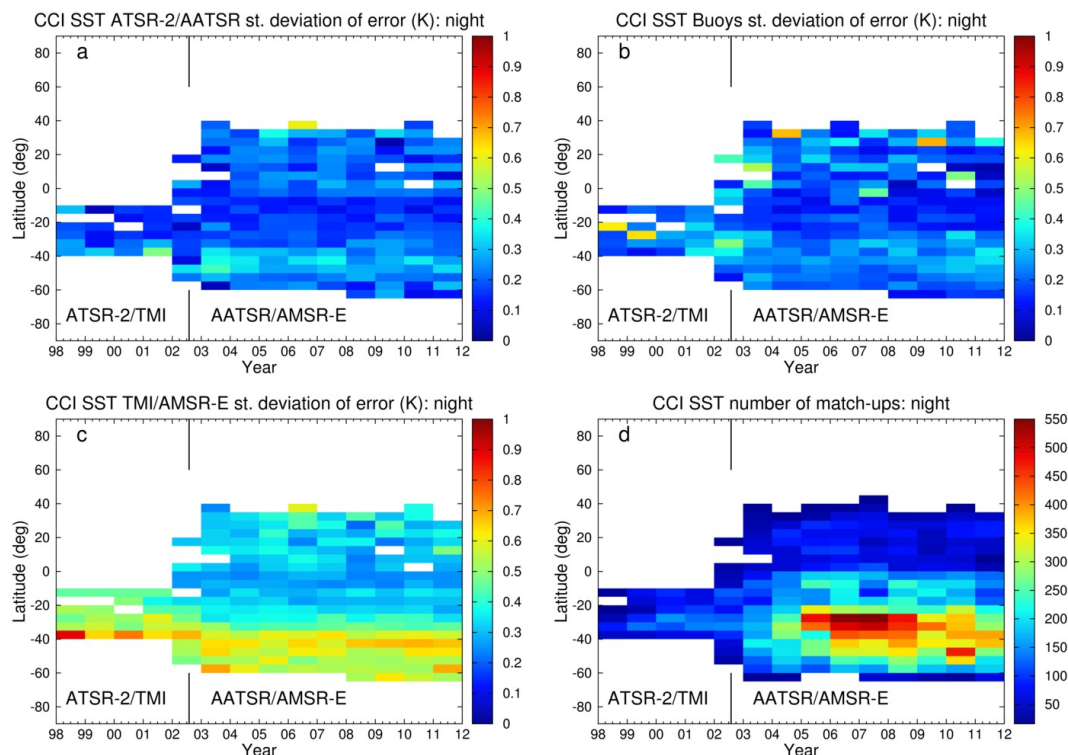


Figure 8. Hovmöller diagrams of triple collocation during night. Standard deviation of the error for (a) ATSR-2/AATSR, (b) drifting buoys and (c) TMI/AMSR-E, while (d) shows the number of collocations. Until end of July 2002, the observations of ATSR-2/TMI have been used and afterwards those of AATSR/AMSR-E. Only grid boxes with at least 20 match-ups are plotted.

Previous studies using the triple collocation approach have found for the drifting buoys standard deviation of the error values of 0.23 K [39], 0.26 K (median value) [29], 0.15–0.19 K [25], 0.20 K [55] and 0.21–0.22 K [56], which are very close to the value of ~ 0.22 K reported here. The respective values found for AATSR are 0.16 K [39], 0.14 K [25] and 0.15–0.30 [56], which are somehow smaller than the value of ~ 0.22 K of this study, but in agreement with Xu and Ignatov [56], who have also applied the triple collocation to the CCI. The results for AMSR-E standard deviation of the error are 0.42 K [39], ~ 0.48 K [25] and 0.28 K [55]. The first two results compare favourably with the value of ~ 0.47 K found here. The much lower value for AMSR-E reported by Gentemann [55] despite using the same dataset version to this study could be a result of averaging the SST from the IR imager (MODIS in Gentemann [55]) to the spatial resolution of the AMSR-E (i.e., 0.25°).

3.5. Validation of Uncertainty

One significant advantage of the CCI dataset is that each SST value is provided with its uncertainty [57,58]. Thus, it is important to assess not only the quality of the SST, but also its associated uncertainty. Figure 9a,b presents how the mean bias and the standard deviation vary against the binned CCI uncertainty for day and night, respectively. Ideally, the bias (red crosses) should be 0 K and the standard deviation (blue squares) should lie on the green line given by the equation $y = \sqrt{0.2^2 + x^2}$, where 0.2 K is the uncertainty of the drifting buoys and x the uncertainty of CCI. Here, the assumption is made that the uncertainty of the drifting buoys is 0.2, which is in line with the above-mentioned results and studies and the review paper by Kennedy [59]. For CCI daytime retrievals, the mean bias increases slightly with increasing CCI uncertainty. Above 0.9 K, the mean bias is oscillating

around 0 K, but there, the number of match-ups is very small (right axis, cyan bars). The number distribution indicates that the majority of daytime CCI has uncertainty values between 0.4 K and 0.6 K. The standard deviation lies rather close, although below to the theoretical line for the uncertainty interval [0.45, 0.9] K, an indication of a slight overestimation of CCI daytime uncertainty. Below the uncertainty level of 0.3 K, the standard deviation is almost constant at 0.4 K, which is higher than the expected value of 0.2 K, thus indicating an underestimation of CCI uncertainty. In total, about 78% of the match-ups are within the combined uncertainty (CCI and drifting buoys), while 94% of them are within twice the combined uncertainty.

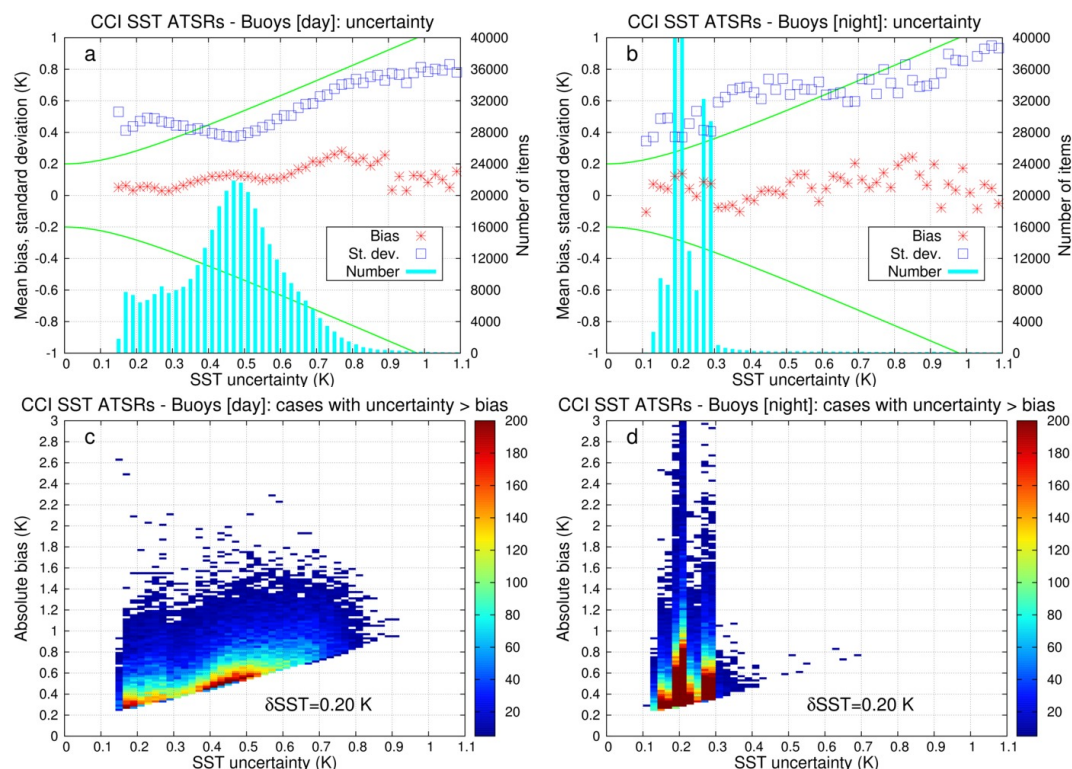


Figure 9. (Top row) Binned mean bias (red crosses) and standard deviation (blue squares) of the difference CCI minus drifting buoys against the CCI uncertainty for (a) day and (b) night. The size of the bin is 0.02 K. The cyan bars indicate the number of match-ups for each bin (right axis). The green line is the theoretical value of the standard deviation for the match-ups by assuming that the standard deviation of the error for the drifting buoys is 0.2 K. (Bottom row) 2D histograms of the absolute bias versus CCI uncertainty for the number of match-ups with SST difference larger than the combined uncertainty for (c) day and (d) night. Again, it is assumed that the standard deviation of the error (δ SST) for the drifting buoys is 0.2 K, and the size of the bin is 0.02 K for both axis. Only grid boxes with at least five pairs having absolute SST difference larger than the combined uncertainty are plotted.

Regarding the CCI night time uncertainty, the situation is more complicated as in general the measurements have an uncertainty of either ~ 0.2 K or ~ 0.28 K (Figure 9b). Only a small number of CCI match-ups have uncertainties in the interval [0.13, 0.3] K, and a very limited number has uncertainty larger than 0.3 K. The bias for the majority of match-ups (having uncertainties of about 0.2 or 0.28 K) is 0.1 K, and the standard deviation is close to the theoretical value, but now mostly indicating an underestimation of the CCI night time uncertainty. The underestimation is more evident for the rest of the match-ups, although not too far away from the theoretical value, while now, the mean bias is closer to 0 K. The poorer performance of the night time uncertainties in comparison to daytime is also reflected in the percentages of the match-ups within the combined uncertainty being 57% and 84% for them within twice the value of the combined uncertainty. Given that uncertainty of

the drifting buoys is the same between day night, this suggests that the CCI night time uncertainty model needs to be improved.

It is useful to further understand to what degree the uncertainty validation of the CCI depends on the uncertainty value of the drifting buoys and the goodness of the CCI uncertainty model. Figure 9c,d presents the two-dimensional histograms of the absolute bias versus the CCI uncertainty for day and night. In these two figures, only pairs with SST difference outside the expected combined CCI and drifting buoys uncertainties are considered. For the daytime conditions, the majority of the pairs is very close to the border, lying almost parallel to it. This means that with a slight increase of either the CCI or the drifting buoys' uncertainty, these pairs would be successful in terms of uncertainty validation. More specifically regarding the CCI, any uncertainty increase should occur only for the uncertainties in the region of 0.2 K. The increased number of match-ups outside the combined uncertainty in the interval 0.4–0.6 K coincides with the bulk of the uncertainty assignment (Figure 9a). It is worth noting that only a limited number of match-ups (less than 40) shows absolute SST differences larger than 0.8 K not captured by the uncertainty model. On the other hand, the underestimation of CCI night time uncertainty is evident from Figure 9d. It is worth considering that the number of match-ups indicating an underestimation of uncertainty is not huge (compare Figure 9b,d), but still, the absolute SST difference reaches up to 0.8 K when the estimated CCI uncertainty is only 0.2 K.

Figure 10 presents the monthly evolution of the percentage for the match-ups that have differences within the combined uncertainty. In order to verify the impact that the uncertainty of drifting buoys can have on these percentages, three values (0.15 K, 0.20 K and 0.25 K) are chosen, which are compatible with literature and this study. During the day, the percentage of match-ups within the uncertainty improves with time from 70–85%, and this is in line with the previously mentioned percentage of 78% for the whole dataset. However, both the periods after the failure of the ATSR-1 3.7 μm channel (May 1992) and during the ERS-2 gyro failure (beginning of 2001) have lower percentages down to 60% indicating that the uncertainty model does not capture these events well. Although it could be argued that the apparent temporal improvement of the percentage is due to drifting buoys, the difference during day among the three percentages (one for each δSST) is too small to justify this (at least for the δSST considered here). For night time retrievals the percentage within the uncertainty shows obvious monthly fluctuations before 2006, while there is some temporal improvement from 60–75%. Surprisingly, it is the first period of ATSR-2 (till the end of 1996) that shows the poorest performance, about 10% lower than ATSR-1. Now, the choice of the drifting buoys uncertainty has a considerable impact on the percentage by increasing it about 10% when δSST increases from 0.15 K–0.25 K. This is expected given that the night time CCI has a similar uncertainty to drifting buoys (Figure 9b).

To summarize, the daytime uncertainty of CCI is better than the respective one during night, but still in both cases the underlying uncertainty model of CCI should be further improved in order to eliminate the temporal evolution as far as possible and to take into account known periods with degraded instrument performance.

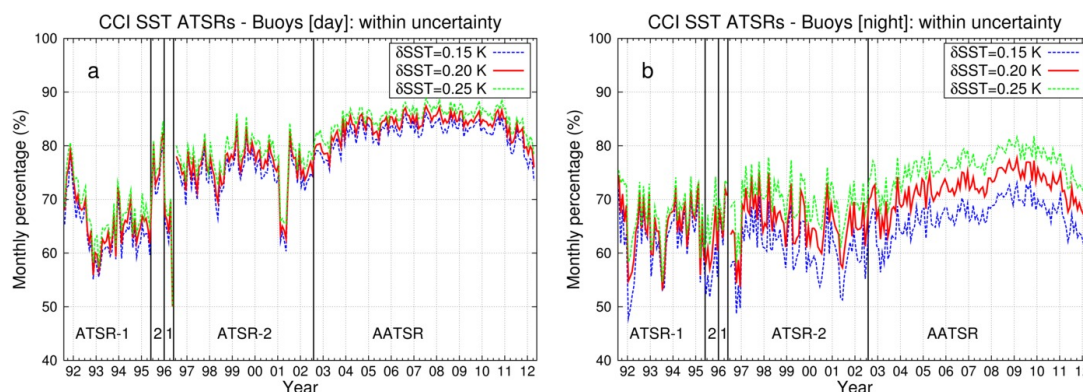


Figure 10. Monthly percentages within combined uncertainty (CCI and drifting buoys) for three different values of uncertainty for the drifting buoys (δSST) during (a) day and (b) night.

4. Comparison with ARC

The ARC SST was the precursor dataset of the CCI, so it is useful to make a comparison between the two. For this purpose, the match-up dataset of Lean and Saunders [25] is used here, which covers the period August 1991–December 2009 (thus no AATSR/drifted buoys collocations for 2010–2012). Both the observations and the collocation approach used by Lean and Saunders [25] are the same as described in Section 2. Similarly to the quality control applied for CCI (Section 2), a ± 3 K threshold is used to eliminate the outliers not removed by the conservative quality control. It is reminded that a three-sigma annual filter has been used in Lean and Saunders [25], while the validation has been performed mainly for 1 m depth SST, which could explain some minor differences between the two studies.

Figure 11 is similar to Figure 3, but presents the comparison between ARC and drifting buoys for the period 1991–2009. The significant warm biases seen in the CCI dataset are absent in ARC both during day and night or at least much less prominent, with ARC being in general in the range $[-0.1, 0.1]$ K from drifting buoys, particularly for night time. The regions appearing consistently with warm SST biases in respect to buoys for both CCI and ARC are off the coast of Indian Peninsula, the Maritime Continent (but not for night time ARC) and the Gulf of Mexico, with the region around the Indian Peninsula being the more challenging (median bias about 0.5 K). During the day, the median bias of CCI is lower than ARC in the Gulf of Mexico and the region of Kuroshio current. There are a few regions with cold biases in ARC, although these are less obvious than CCI with values close to -0.1 K. In terms of robust standard deviation, again the ARC performs better than CCI, and especially during night, the RSD is almost everywhere less than 0.2 K.

However, the direct comparison between Figures 3 and 11 is not really valid as they do not cover exactly the same period, despite the fact that AATSR CCI was pretty stable for its whole lifetime (Section 3). For this reason, Table 3 presents the statistics of the comparison between both CCI and ARC with drifting buoys. The results for all statistical measures of CCI are very similar (maximum difference ± 0.03 K) between Tables 1 and 3, despite that the period examined in the Tables is different for all three ATSRs. This shows the robustness of the overall statistics to small changes of the time period considered for CCI. Returning to the comparison between CCI and ARC, it can be seen that the number of match-ups is slightly different between the two (Table 3) with ARC having in general more match-ups. This is the outcome of different spatial resolutions 0.05° for CCI vs. 0.1° for ARC. All statistical parameters show better agreement of ARC with drifting buoys than CCI. Nevertheless, the differences are not large and in general are within ± 0.05 K. However, the difference in the statistics between CCI and ARC for ATSR-1 is puzzling, especially regarding the biases. The SST retrieval method is the same between the two datasets, and the only difference is the spatial resolution.

In order to investigate the temporal evolution of quality between CCI and ARC, Figure 12 presents the annual median bias and robust standard deviation. The better performance of ARC than CCI in terms of median bias (being closer to 0 K) can be seen for the whole period from 1991–2009. However, CCI is more consistent regarding the differences between day and night for the periods of ATSR-1 and AATSR. During the ATSR-2 period (1997–2000), the performance of CCI is almost identical to ARC for the daytime retrieval. The sign of the median bias is not the same before 2000 during night, with CCI presenting a warm bias and ARC having a cold bias. The difference in performance between CCI and ARC during night reaches about 0.25 K in 1993–1994 (Figure 12a). As mentioned previously, this big difference is surprising as both CCI and ARC use practically the same cloud mask and SST retrieval method in this period. Concerning the time evolution of the RSD, this is very similar between CCI and ARC, with CCI having larger values by ~ 0.05 K from ARC. It is interesting to note that ARC daytime is not impacted by the switch from ATSR-1 to ATSR-2 in 1996, while the CCI night time is not impacted by the gyro failure on ERS-2 in 2001 (in accordance with Figure 5). Furthermore, the daytime ARC RSD is very similar to the night time RSD of CCI.

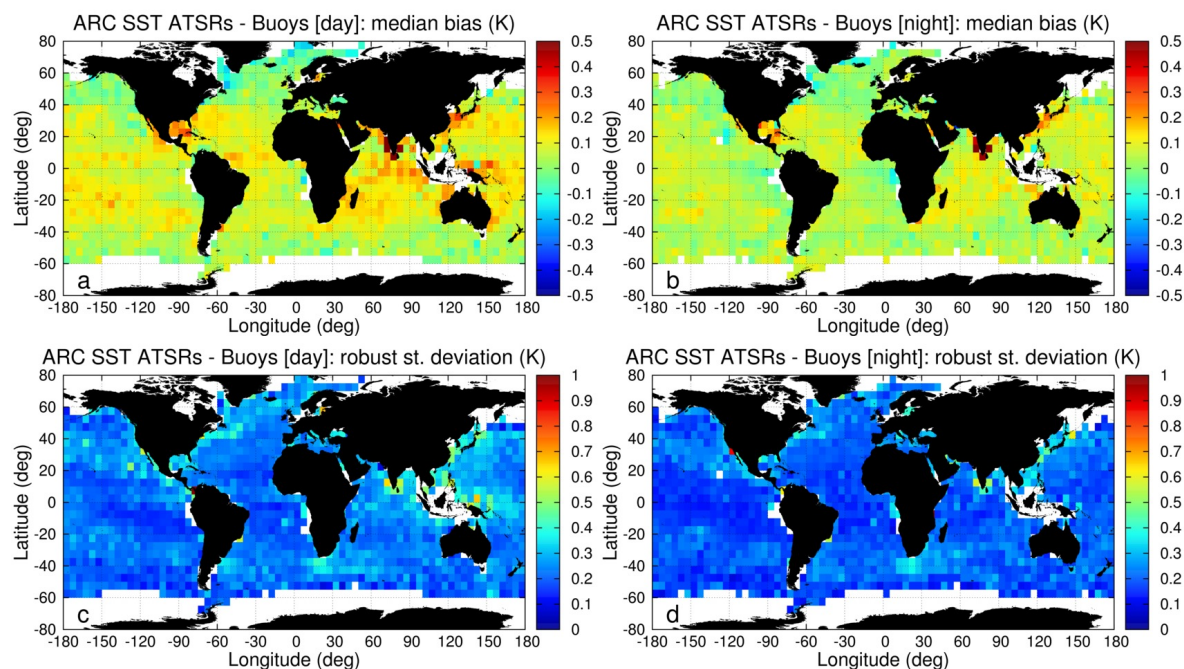


Figure 11. The median bias (a,b) and robust standard deviation (c,d) of ARC SST minus drifting buoys averaged in $5^\circ \times 5^\circ$ boxes for the period: August 1991–December 2009. The results for daytime (night time) observations are presented in left (right) column. Gridboxes with fewer than 20 match-ups appear white.

Table 3. Same as Table 1. Statistics are provided both for CCI and ARC against drifting buoys for the period indicated in the parentheses.

Algorithm	Mean Bias (K)	St. Deviation (K)	Median Bias (K)	RSD (K)	Number
CCI					
<i>ATSR-1 (08/1991–05/1995)</i>					
Day	0.15	0.65	0.14	0.53	11,060
Night	0.13	0.67	0.14	0.53	9373
<i>ATSR-2 (07/1996–06/2002)</i>					
Day	0.10	0.53	0.10	0.39	50,125
Night	0.10	0.46	0.11	0.28	33,605
<i>AATSR (08/2002–12/2009)</i>					
Day	0.11	0.43	0.11	0.28	194,786
Night	0.11	0.38	0.11	0.22	152,962
ARC					
<i>ATSR-1 (08/1991–05/1995)</i>					
Day	0.01	0.61	0.02	0.46	13,963
Night	−0.09	0.58	−0.07	0.42	10,779
<i>ATSR-2 (07/1996–06/2002)</i>					
Day	0.08	0.48	0.08	0.31	50,639
Night	0.04	0.43	0.04	0.24	37,333
<i>AATSR (08/2002–12/2009)</i>					
Day	0.08	0.37	0.09	0.21	185,114
Night	0.07	0.34	0.06	0.18	156,662

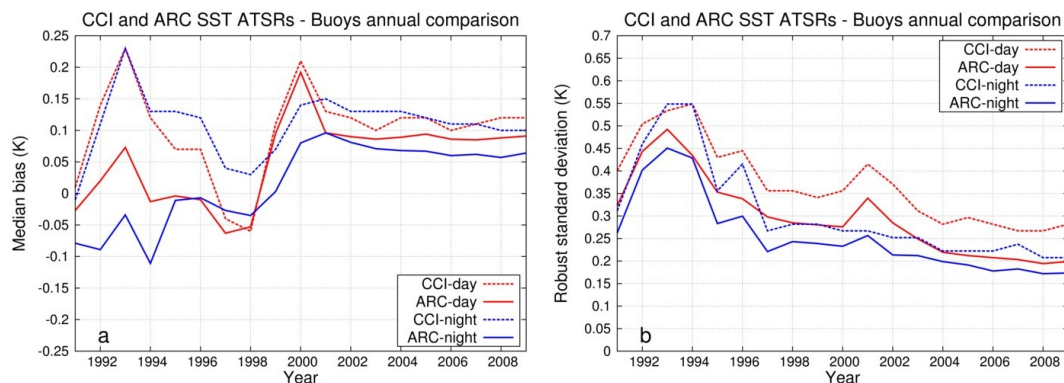


Figure 12. Annual (a) median bias and (b) robust standard deviation of CCI (dotted lines) and ARC (solid lines) against drifting buoys. The results are given for day (red lines) and night (blue lines) during 1991 to 2009.

Figure 13 shows the monthly evolution of percentages for four ad hoc thresholds (the same as Figure 6), but for ARC. In this way, performance changes at sub-annual time scale are more easily discerned, thus providing complementary information to Figure 12. The superiority of ARC is confirmed with percentages of collocations inside ± 0.1 K or ± 0.2 K being about 10% higher than CCI. It is worth noting that for night time ARC the percentage within 0.2 K reaches 70% after 2005. Regarding the percentages with absolute differences larger than 2 K, there is no difference between CCI and ARC, although ARC shows smoother evolution. For the percentage of $|\Delta SST|$ larger than 1 K, ARC displays lower values than CCI, especially during the ATSR-1 period by $\sim 5\%$. The failure of ATSR-1 3.7 μm channel (May 1992) and the ERS-2 gyro (beginning of 2001) are clearly seen in the time series, together with the switch from ATSR-1 to ATSR-2. Although, in ARC, the first months of ATSR-2 have higher percentages (Figure 13a) than the respective period of ATSR-1 (before the 3.7 μm channel failure), as CCI the two instruments seem to have a similar quality.

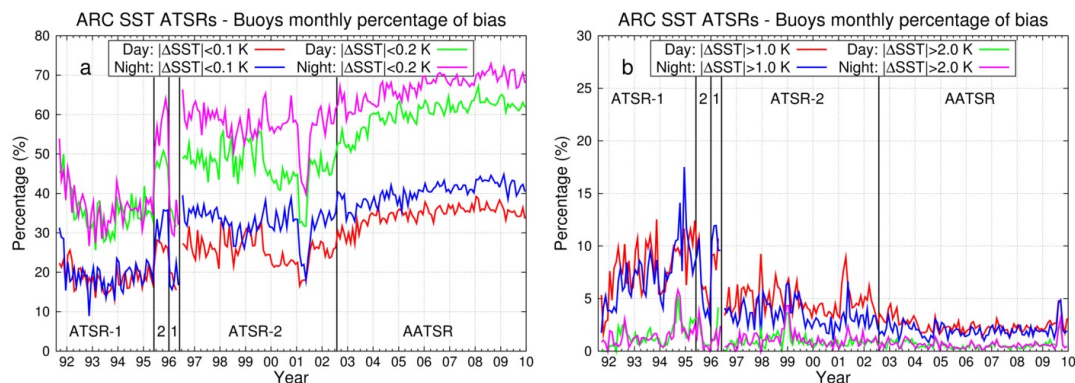


Figure 13. Same as Figure 6, but for ARC during August 1991 to December 2009.

5. Summary and Conclusions

The CCI SSTs (Phase 1, Version 1.1) from the ATSRs covering the period 1991–2012 have been assessed against collocated observations from drifting buoys and also compared with the corresponding comparisons for the ARC (the precursor dataset of CCI SST). Only the performance of CCI SSTs from the ATSRs has been assessed here. For the comparison of the CCI and ARC datasets, it should be noted that apart from ATSR-1, the retrieval methods for the skin SSTs are different. Furthermore, the ARC dataset is provided at a spatial resolution of 0.1° , while CCI has a finer resolution of 0.05° . Because of this, the CCI standard deviation is expected to be worse than ARC due to higher random and sampling uncertainties [57,58,60].

The median bias of CCI is slightly larger than 0.1 K and the robust standard deviation (RSD) in the range 0.22–0.55 K lower for night time than daytime conditions, and this improves with instrument generation (AATSR has the smallest and ATSR-1 the largest) (Table 1). However, considering overlapping periods for the instruments indicates that the performance of ATSR-2 is similar to AATSR (Table 2). The median bias is larger in the tropics ~ 0.3 K during the day, with lower values (0.1–0.2 K) and more homogeneous distribution during the night. Some regions show negative biases, e.g., due to the dust outflow from the Sahara and Arabian deserts for night time conditions. The Maritime Continent has a peak for daytime RSD at about 0.6 K. The median bias is fairly stable (0.1–0.15 K) since 2000–2001, but variable before then. The comparable performance of ATSR-2 and AATSR is demonstrated in the time series of the RSD being 0.2–0.3 K during night and about 0.1 K higher during the day. On the other hand, ATSR-1 shows increasing RSD with time from 0.3 K to over 0.6 K. Forty to sixty percent of the match-ups have absolute SST differences less than 0.2 K (higher percentages during night), except for the periods after the failure of the $3.7\ \mu\text{m}$ channel on ATSR-1 and the gyro failure on ERS-2. Similarly, less than 10% of match-ups show absolute SST differences larger than 1 K (again except for the above mentioned periods), decreasing to 5% for the AATSR period.

Triple collocation analysis has been used for the first time, to the best of our knowledge, for the period 1998–2002 using TMI and ATSR-2 observations, extending backwards in time the estimation of the random uncertainty of SST from drifting buoys, IR imagers and MW imagers. The application of triple collocation under daytime conditions when the wind speed is larger than 10 m/s provides meaningful results for the period of AATSR/AMSR-E. Both day and night results indicate that, since 2004, drifting buoys and CCI AATSR have a random uncertainty of about 0.22 K, stable with time. Before 2004, drifting buoys have larger values (~ 0.3 K), while ATSR-2 shows slightly lower values (~ 0.2 K). The random uncertainty for AMSR-E is about 0.47 K, stable with time, while TMI has as expected higher values of ~ 0.55 K. The Hovmöller diagrams of the triple collocation during night indicate that the analysis samples mainly the Southern Hemisphere mid-latitudes, producing mostly latitudinally homogeneous results (in addition to time) for the drifting buoys and the ATSR-2/AATSR CCI SST. Surprisingly, the performance of AMSR-E is much better in the tropics (~ 0.3 K), almost half of the value found in the mid-latitudes, which is comparable to TMI. This is the first time to the best of our knowledge that the AMSR-E SST random uncertainty has been reported to change with latitude.

The CCI provides uncertainty for each SST retrieval, and its validation showed that the model used during the daytime slightly overestimates uncertainty in the interval [0.45, 0.9] K and underestimates below 0.3 K. Small underestimations of the uncertainty are also seen for the night time conditions, but now for some of the match-ups, the uncertainties provided are too low. The time evolution of the percentage of match-ups within combined uncertainty (drifting buoys and CCI) is 70–80% for daytime, though the model does not capture the instrument failure periods of the ATSR-1 $3.7\ \mu\text{m}$ channel and the ERS-2 gyro. The corresponding percentages for night are lower, being 60–70%. However, the night percentages depend on the assigned value of uncertainty for the drifting buoys being at the same level, which is not the case during the day.

Results show that CCI (Phase 1) does not yet match the quality of ARC SST retrievals when comparing to drifting buoys. The value of the ARC median bias is closer to 0 K than CCI, and the RSD is about 0.05 K lower for ARC. This means that the night time CCI RSD more or less matches the ARC RSD during the day. Although the results of the comparison against drifting buoys are not very different, the superiority of ARC can also be seen in the more homogeneous geographical distribution of median bias and RSD than CCI. However, there are regions (e.g., Gulf of Mexico, Kuroshio) where CCI is better than ARC, while the region around the Indian Peninsula has proven difficult for both datasets. The same picture emerges for the temporal evolution of the percentages of $|\Delta\text{SST}|$ within 0.1/0.2 K or larger than 1 K with ARC manifesting better performance than CCI. This fact probably indicates that the SST retrieval approach based on coefficients estimated from radiative transfer simulations performs better than the optimal estimation for dual view observations; even if for single view observations, the optimal estimation retrieval provides better results [26]. An alternative possibility is that the optimal

estimation retrieval has not been implemented correctly in CCI, e.g., by not removing all the subtle biases of the radiances. Nevertheless, the observed discrepancies between CCI and ARC during the period of ATSR-1 are unexplained given that both datasets use the same retrieval method; especially concerning bias, as the standard deviation of CCI is expected to be higher than ARC due to the finer spatial resolution.

It should be noted that although the CCI (Phase 1) SSTs from the ATSRs are not as good as ARC, they are in line or even better than other SST datasets based on thermal infrared (TIR) observations [30–32,61]. The CCI also includes SSTs from the Advanced Very High Resolution Radiometers (AVHRRs) and an SST analysis using both ATSR and AVHRR [22]. The combination of the high quality SSTs provided by the ATSRs and the global coverage provided by AVHRRs offers an attractive dataset for climate studies. Indeed, the CCI SST dataset is already being used successfully [56,62] for climate research. Some of the shortcomings in the CCI SSTs identified here are being addressed in Phase 2 of the CCI, while other enhancements could be incorporated in future versions [63].

Acknowledgments: TMI and AMSR-E data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science MEaSUREs DISCOVER Project. Data are available at the www.remss.com. Both CCI and ARC SST datasets can be downloaded from CEDA <http://catalogue.ceda.ac.uk>. Christoforos Tsamalis was supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). Roger Saunders has been partially funded by the ESA CCI CMUG project (RFQ/3-12943/09/I-OL).

Author Contributions: C.T. designed and performed the analysis and wrote the paper. R.S. contributed in the analysis and the writing of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AATSR	Advanced Along Track Scanning Radiometer
AIRS	Atmospheric InfraRed Sounder
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
ARC	ATSR Reprocessing for Climate
ATSR	Along Track Scanning Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CCI	Climate Change Initiative
Δ SST	Difference of SST
δ SST	Uncertainty of SST
ECV	Essential Climate Variable
ENVISAT	ENVironmental SATellite
EOS	Earth Observing System
ERS	European Remote Sensing
ESA	European Space Agency
GSICS	Global Space-based Inter-Calibration System
GTS	Global Telecommunication System
IASI	Infrared Atmospheric Sounding Interferometer
ICODS	International Comprehensive Ocean-Atmosphere Data Set
IR	InfraRed
MetOp	Meteorological Operational
MODIS	MODerate resolution Imaging Spectroradiometer
NIR	Near InfraRed
NWP	Numerical Weather Prediction
RSD	Robust Standard Deviation
SST	Sea Surface Temperature
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measuring Mission
MW	MicroWave

References

- Kennedy, J.J.; Rayner, N.A.; Smith, R.O.; Parker, D.E.; Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.* **2011**, *116*, D14103.
- Kennedy, J.J.; Rayner, N.A.; Smith, R.O.; Parker, D.E.; Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.* **2011**, *116*, D14104.
- Emery, W.J.; Castro, S.; Wick, G.A.; Schluessel, P.; Donlon, C. Estimating sea surface temperature from infrared satellite and in situ temperature data. *Bull. Am. Meteorol. Soc.* **2001**, *82*, 2773–2785.
- Donlon, C.; Rayner, N.; Robinson, I.; Poulter, D.J.S.; Casey, K.S.; Vazquez-Cuervo, J.; Armstrong, E.; Bingham, A.; Arino, O.; Gentemann, C.; et al. The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 1197–1213.
- Deschamps, P.Y.; Phulpin, T. Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 μm . *Bound. Lay. Meteorol.* **1980**, *18*, 131–143.
- Casey, K.S.; Brandon, T.B.; Cornillon, P.; Evans, R. The past, present and future of the AVHRR Pathfinder SST program. In *Oceanography from Space: Revisited*; Barale, V., Gower, J.F.R., Alberotanza, L., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 273–287.
- Merchant, C.J.; Embury, O.; Rayner, N.A.; Berry, D.I.; Corlett, G.K.; Lean, K.; Veal, K.L.; Kent, E.C.; Llewellyn-Jones, D.T.; Remedios, J.J.; et al. A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. *J. Geophys. Res.* **2012**, *117*, C12013.
- 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.; et al. The accuracy of SST retrievals from AATSR: An initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets. *Adv. Space Res.* **2006**, *37*, 764–769.
- Kennedy, J.J.; Smith, R.O.; Rayner, N.A. Using AATSR data to assess the quality of in situ sea-surface temperature observations for climate studies. *Remote Sens. Environ.* **2012**, *116*, 79–92.
- Atkinson, C.P.; Rayner, N.A.; Roberts-Jones, J.; Smith, R.O. Assessing the quality of sea surface temperature observations from drifting buoys and ships on a platform-by-platform basis. *J. Geophys. Res.* **2013**, *118*, 3507–3529.
- Blackmore, T.; O’Carroll, A.; Fennig, K.; Saunders, R. Correction of AVHRR Pathfinder SST data for volcanic aerosol effects using ATSR SSTs and TOMS aerosol optical depth. *Remote Sens. Environ.* **2012**, *116*, 107–117.
- Le Borgne, P.; Marsouin, A.; Orain, F.; Roquet, H. Operational sea surface temperature bias adjustment using AATSR data. *Remote Sens. Environ.* **2012**, *116*, 93–106.
- Roberts-Jones, J.; Fiedler, E.K.; Martin, M.J. Daily, global, high-resolution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. *J. Clim.* **2012**, *25*, 6215–6232.
- Roberts-Jones, J.; Bovis, K.; Martin, M.J.; McLaren, A. Estimating background error covariance parameters and assessing their impact in the OSTIA system. *Remote Sens. Environ.* **2016**, *176*, 117–138.
- Llewellyn-Jones, D.; Edwards, M.C.; Mutlow, C.T.; Birks, A.R.; Barton, I.J.; Tait, H. AATSR: Global-change and surface-temperature measurements from Envisat. *ESA Bull.* **2001**, *105*, 11–21.
- Noyes, E.J.; Minnett, P.J.; Remedios, J.J.; Corlett, G.K.; Good, S.A.; Llewellyn-Jones, D.T. The accuracy of the AATSR sea surface temperatures in the Caribbean. *Remote Sens. Environ.* **2006**, *101*, 38–51.
- O’Carroll, A.G.; Saunders, R.W.; Watts, J.G. The measurement of the sea surface temperature by satellites from 1991 to 2005. *J. Atmos. Ocean. Technol.* **2006**, *23*, 1573–1582.
- Reynolds, R.W.; Gentemann, C.L.; Corlett, G.K. Evaluation of AATSR and TMI satellite SST data. *J. Clim.* **2010**, *23*, 152–165.
- Smith, D.; Mutlow, C.; Delderfield, J.; Watkins, B.; Mason, G. ATSR infrared radiometric calibration and in-orbit performance. *Remote Sens. Environ.* **2012**, *116*, 4–16.
- Illingworth, S.M.; Remedios, J.J.; Parker, R.J. Intercomparison of integrated IASI and AATSR calibrated radiances at 11 and 12 μm . *Atmos. Chem. Phys.* **2009**, *9*, 6677–6683.
- Bali, M.; Mittaz, J.P.; Maturi, E.; Goldberg, M.D. Comparisons of IASI-A and AATSR measurements of top-of-atmosphere radiance over an extended period. *Atmos. Meas. Tech.* **2016**, *9*, 3325–3336.

22. Merchant, C.J.; Embury, O.; Roberts-Jones, J.; Fiedler, E.; Bulgin, C.E.; Corlett, G.K.; Good, S.; McLaren, A.; Rayner, N.; Morak-Bozzo, S.; et al. Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). *Geosci. Data J.* **2014**, *1*, 179–191.
23. Hollmann, R.; Merchant, C.J.; Saunders, R.; Downy, C.; Buchwitz, M.; Cazenave, A.; Chuvieco, E.; Defourny, P.; de Leeuw, G.; Forsberg, R.; et al. The ESA Climate Change Initiative: Satellite data records for essential climate variables. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1541–1552.
24. Merchant, C.J.; Llewellyn-Jones, D.; Saunders, R.W.; Rayner, N.A.; Kent, E.C.; Old, C.P.; Berry, D.; Birks, A.R.; Blackmore, T.; Corlett, G.K.; et al. Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Adv. Space Res.* **2008**, *41*, 1–11.
25. Lean, K.; Saunders, R.W. Validation of the ATSR Reprocessing for Climate (ARC) dataset using data from drifting buoys and a three-way error analysis. *J. Clim.* **2013**, *26*, 4758–4772.
26. Merchant, C.J.; Le Borgne, P.; Marsouin, A.; Roquet, H. Optimal estimation of sea surface temperature from split-window observations. *Remote Sens. Environ.* **2008**, *112*, 2469–2484.
27. Embury, O.; Merchant, C.J. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. *Remote Sens. Environ.* **2012**, *116*, 47–61.
28. Embury, O.; Merchant, C.J.; Filipiak, M.J. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Basis in radiative transfer. *Remote Sens. Environ.* **2012**, *116*, 32–46.
29. Xu, F.; Ignatov, A. Evaluation of in situ sea surface temperatures for use in the calibration and validation of satellite retrievals. *J. Geophys. Res.* **2010**, *115*, C09022.
30. Petrenko, B.; Ignatov, A.; Kihai, Y.; Stroup, J.; Dash, P. Evaluation and selection of SST regression algorithms for JPSS VIIRS. *J. Geophys. Res.* **2014**, *119*, 4580–4599.
31. Kilpatrick, K.A.; Podesta, G.; Walsh, S.; Williams, E.; Halliwell, V.; Szczodrak, M.; Brown, O.B.; Minnett, P.J.; Evans, R. A decade of sea surface temperature from MODIS. *Remote Sens. Environ.* **2015**, *165*, 27–41.
32. Marsouin, A.; Le Borgne, P.; Legendre, G.; Pere, S.; Roquet, H. Six years of OSI-SAF METOP-A AVHRR sea surface temperature. *Remote Sens. Environ.* **2015**, *159*, 288–306.
33. Merchant, C.J.; Harris, A.R.; Maturi, E.; MacCallum, S. Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 2765–2755.
34. Merchant, C.J.; Embury, O.; Le Borgne, P.; Bellec, B. Saharan dust in nighttime thermal imagery: Detection and reduction of related biases in retrieved sea surface temperature. *Remote Sens. Environ.* **2006**, *104*, 15–30.
35. Good, E.J.; Kong, X.; Embury, O.; Merchant, C.J.; Remedios, J. An infrared desert dust index for the Along-Track Scanning Radiometers. *Remote Sens. Environ.* **2012**, *116*, 159–176.
36. Horrocks, L.A.; Candy, B.; Nightingale, T.J.; Saunders, R.W.; O’Carroll, A.; Harris, A.R. Parameterizations of the ocean skin effect and implications for satellite-based measurement of sea-surface temperature. *J. Geophys. Res.* **2003**, *108*, 3096.
37. Donlon, C.J.; Minnett, P.J.; Gentemann, C.; Nightingale, T.J.; Barton, I.J.; Ward, B.; Murray, M.J. Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Clim.* **2002**, *15*, 353–369.
38. Veal, K.L.; Corlett, G.K.; Ghent, D.; Llewellyn-Jones, D.T.; Remedios, J.J. A time series of mean global skin SST anomaly using data from ATSR-2 and AATSR. *Remote Sens. Environ.* **2013**, *135*, 64–76.
39. O’Carroll, A.G.; Eyre, J.R.; Saunders, R.W. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *J. Atmos. Ocean. Technol.* **2008**, *25*, 1197–1207.
40. Zwieback, S.; Scipal, K.; Dorigo, W.; Wagner, W. Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Process. Geophys.* **2012**, *19*, 69–80.
41. Wentz, F.J.; Gentemann, C.; Smith, D.; Chelton, D. Satellite measurements of sea surface temperature through clouds. *Science* **2000**, *288*, 847–850.
42. Chelton, D.B.; Wentz, F. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 1097–1115.
43. Gentemann, C.L.; Meissner, T.; Wentz, F.J. Accuracy of satellite sea surface temperatures at 7 and 11 GHz. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1009–1017.
44. Woodruff, S.D.; Worley, S.J.; Lubker, S.J.; Ji, Z.; Freeman, J.E.; Berry, D.I.; Brohan, P.; Kent, E.C.; Reynolds, R.W.; Smith, S.R.; et al. ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **2011**, *31*, 951–967.

45. O'Carroll, A.G.; Watts, J.G.; Horrocks, L.A.; Saunders, R.W.; Rayner, N.A. Validation of the AATSR meteo product sea surface temperature. *J. Atmos. Ocean. Technol.* **2006**, *23*, 711–726.
46. Atkinson, C.P.; Rayner, N.A.; Kennedy, J.J.; Good, S.A. An integrated database of ocean temperature and salinity observations. *J. Geophys. Res.* **2014**, *119*, 7139–7163.
47. Xu, F.; Ignatov, A. In situ SST Quality Monitor (iQuam). *J. Atmos. Ocean. Technol.* **2014**, *31*, 164–180.
48. Peyridieu, S.; Chédin, A.; Capelle, V.; Tsamalis, C.; Pierangelo, C.; Armante, R.; Crevoisier, C.; Crépeau, L.; Siméon, M.; Ducos, F.; et al. Characterisation of dust aerosols in the infrared from IASI and comparison with PARASOL, MODIS, MISR, CALIOP, and AERONET observations. *Atmos. Chem. Phys.* **2013**, *13*, 6065–6082.
49. Tsamalis, C.; Chédin, A.; Pelon, J.; Capelle, V. The seasonal vertical distribution of the Saharan Air Layer and its modulation by the wind. *Atmos. Chem. Phys.* **2013**, *13*, 11235–11257.
50. Gentemann, C.L.; Donlon, C.J.; Stuart-Menteth, A.; Wentz, F.J. Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.* **2003**, *30*, 1140.
51. Gentemann, C.L.; Minnett, P.J.; Le Borgne, P.; Merchant, C.J. Multi-satellite measurements of large diurnal warming events. *Geophys. Res. Lett.* **2008**, *35*, L22602.
52. Merchant, C.J.; Filipiak, M.J.; Borgne, P.L.; Roquet, H.; Autret, E.; Piolle, J.; Lavender, S. Diurnal warm-layer events in the western Mediterranean and European shelf seas. *Geophys. Res. Lett.* **2008**, *35*, L04601.
53. Gentemann, C.L.; Wentz, F.J.; Mears, C.A.; Smith, D.K. In situ validation of Tropical Rainfall Measuring Mission microwave sea surface temperatures. *J. Geophys. Res.* **2004**, *104*, C04021.
54. Morak-Bozzo, S.; Merchant, C.J.; Kent, E.C.; Berry, D.I.; Carella, G. Climatological diurnal variability in sea surface temperature characterized from drifting buoy data. *Geosci. Data J.* **2016**, *3*, 20–28.
55. Gentemann, C.L. Three way validation of MODIS and AMSR-E sea surface temperatures. *J. Geophys. Res.* **2014**, *119*, 2583–2598.
56. Xu, F.; Ignatov, A. Error characterization in iQuam SSTs using triple collocations with satellite measurements. *Geophys. Res. Lett.* **2016**, *43*, 1–9.
57. Bulgin, C.E.; Embury, O.; Corlett, G.; Merchant, C.J. Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments. *Remote Sens. Environ.* **2016**, *178*, 213–222.
58. Bulgin, C.E.; Embury, O.; Merchant, C.J. Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data. *Remote Sens. Environ.* **2016**, *177*, 287–294.
59. Kennedy, J.J. A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* **2014**, *52*, 1–32.
60. Liu, Y.; Minnett, P.J. Sampling errors in satellite-derived infrared sea-surface temperatures. Part I: Global and regional MODIS fields. *Remote Sens. Environ.* **2016**, *177*, 48–64.
61. O'Carroll, A.G.; August, T.; Le Borgne, P.; Marsouin, A. The accuracy of SST retrievals from Metop-A IASI and AVHRR using the EUMETSAT OSI-SAF matchup dataset. *Remote Sens. Environ.* **2012**, *126*, 184–194.
62. Massonnet, F.; Bellprat, O.; Guemas, V.; Doblas-Reyes, F.J. Using climate models to estimate the quality of global observational data sets. *Science* **2016**, *354*, 452–455.
63. Bulgin, C.E.; Eastwood, S.; Embury, O.; Merchant, C.J.; Donlon, C. The sea surface temperature climate change initiative: Alternative image classification algorithms for sea-ice affected oceans. *Remote Sens. Environ.* **2015**, *162*, 396–407.

