



Norsk institutt for luftforskning
Norwegian Institute for Air Research

Land Surface Temperature Validation for WACMOS-ET

Reference Input Data Set Validation Report

Philipp Schneider



NILU report 9/2017

Contents

Contents	3
List of Figures	5
List of Tables	8
Summary	9
1 Introduction	10
2 Data and Methodology	10
2.1 LST retrieval	10
2.2 In situ data	12
2.2.1 Dedicated in situ validation sites for LST	12
2.2.2 SURFRAD sites	13
2.2.3 ARM sites	15
2.3 Independent remote sensing data	15
2.4 Processing	17
2.4.1 Processing of in situ data	18
2.4.2 Processing of independent remote sensing data	19
3 Validation of AATSR	20
3.1 Validation with in situ observations	20
3.2 Validation against independent EO data	25
3.2.1 Comparison with MODIS MOD11	26
3.2.2 Comparison with GlobTemperature	34
4 Validation of MTSAT	39
4.1 Validation with in situ observations	39
4.2 MTSAT Validation with EO data	40
4.2.1 Inter-comparison of full moasic	41
4.2.2 Tile-by-tile inter-comparison of full global moasic	45
4.2.3 Dependence on zenith view angle	48
4.2.4 Dependence on land cover	52
5 Validation of GOES	58
5.1 Validation against in situ data	58
5.2 Validation against independent EO data	62
6 Validation of SEVIRI	65
7 Conclusions	66
Bibliography	68
Appendices	72
A Global maps of AATSR	72
B Global maps of MODIS	74

C	Maps of MTSAT	76
D	Maps of GOES-E	80

List of Figures

1	Overview of the various LST validation categories and the corresponding accuracy classes as described in (Schneider et al., 2012).	11
2	The Gobabeb LST validation site operated by the Karlsruhe Institute of Technology (Göttsche et al., 2013)	13
3	Geographic overview and bathymetry of the Lake Tahoe Study Site.	14
4	One of the four Lake Tahoe buoys measuring LST as well as a variety of meteorological parameters.	14
5	Map of the location of the SURFRAD sites within the contiguous United States.	15
6	Example of a SURFRAD station, here showing the Desert Rock (Nevada) station.	16
7	Example of one tile of the MOD11A1 LST product, here showing the nighttime LST data for the tile h19v10 along the border between Namibia and Angola.	17
8	Scatter plots of original AATSR-derived nighttime LST against LST computed from observations at in situ stations.	20
9	Scatter plots of original AATSR-derived daytime LST against LST computed from observations at in situ stations.	22
10	Scatter plots of AATSR-derived nighttime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.	23
11	Scatter plots of AATSR-derived daytime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.	24
12	Inter-comparison of the full nighttime (top) and daytime (bottom) LST time series derived from AATSR versus MODIS-Terra for the entire WACMOS-ET study period at the Bondville station.	27
13	Inter-comparison of the full nighttime (top) and daytime (bottom) LST time series derived from AATSR versus MODIS-Terra for the entire WACMOS-ET study period at the Table Mountain station.	28
14	Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for daytime data for 15 July 2007.	29
15	Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for nighttime data for 15 July 2007.	30
16	Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for daytime data for 15 December 2007.	30
17	Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for nighttime data for 15 December 2007.	31
18	Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for daytime data for 15 July 2007. Yellow markers indicate the median value for various temperature classes.	31
19	Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for nighttime data for 15 July 2007. Yellow markers indicate the median value for various temperature classes.	32
20	Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for daytime data for 15 December 2007. Yellow markers indicate the median value for various temperature classes.	32
21	Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for nighttime data for 15 December 2007. Yellow markers indicate the median value for various temperature classes.	33

22	Comparison of the GlobTemperature AATSR LST product versus the WACMOS-ET derived AATSR LST product, shown here for the year 2007 and nine locations with in situ observations.	35
23	Differences between the GlobTemperature AATSR LST product and the WACMOS-ET derived AATSR LST product against the corresponding in situ LST observations, shown here for the year 2007 at nine locations.	36
24	As Figure 22 but using the original L1 cloud mask for the GlobTemperature product.	37
25	As Figure 23 but using the original L1 cloud mask for the GlobTemperature product.	38
26	Time series of in situ LST observations at the Darwin station and MTSAT retrievals at a nearby grid cell.	39
27	Scatterplot of LST ground observations at Darwin, Australia, against LST retrieved from MTSAT over a nearby location.	40
28	Similar to Figure 26 but only showing two days in July 2007.	41
29	Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for daytime data on 15 July 2007.	42
30	Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for nighttime data on 15 July 2007.	43
31	Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for daytime data on 15 December 2007.	43
32	Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for nighttime data on 15 December 2007.	44
33	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 July 2007.	44
34	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 July 2007.	45
35	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 December 2007.	45
36	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 December 2007.	46
37	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 July 2007.	50
38	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 July 2007.	50
39	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 December 2007.	51
40	Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 December 2007.	51
41	Global map of land cover for 2007 as provided by the MODIS MCD12Q1 product.	52
42	Comparison of mean error statistics averaged for both daytime and nighttime data and the two test dates	55
43	Scatter plots of original GOES-derived daytime LST against LST computed from observations at in situ stations.	58
44	Scatter plots of original GOES-derived nighttime LST against LST computed from observations at in situ stations.	59
45	Scatter plots of GOES-derived daytime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.	60
46	Scatter plots of GOES-derived nighttime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.	61
47	Overall scatterplot including summary statistics of all matchups between cloud-corrected GOES LST and in situ LST at all stations and all times. Yellow markers indicate the median GOES LST for various classes of in situ LST.	62

48	Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and GOES-derived LST for daytime data for 15 July 2007. The daytime MODIS mosaic was matched with GOES-E data observed at 14:45 UTC.	63
49	Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and GOES-derived LST for nighttime data for 15 July 2007. The nighttime MODIS mosaic was matched with GOES-E data observed at 02:45 UTC.	63
50	LST from the Gobabeb station versus MSG/SEVIRI LST from the Land-SAF for 10-16 March 2010 (from Göttsche et al. (2013)).	65
51	Global mosaic of the AATSR product on 15 July 2007 for daytime data.	72
52	Global mosaic of the AATSR product on 15 July 2007 for nighttime data.	72
53	Global mosaic of the AATSR product on 15 December 2007 for daytime data.	73
54	Global mosaic of the AATSR product on 15 December 2007 for nighttime data.	73
55	Global mosaic of the MODIS MOD11A1 product on 15 July 2007 for daytime data.	74
56	Global mosaic of the MODIS MOD11A1 product on 15 July 2007 for nighttime data.	74
57	Global mosaic of the MODIS MOD11A1 product on 15 December 2007 for daytime data.	75
58	Global mosaic of the MODIS MOD11A1 product on 15 December 2007 for nighttime data.	75
59	Mosaic of the MTSAT LST product on 15 July 2007 for daytime data.	76
60	Mosaic of the MTSAT LST product on 15 July 2007 for nighttime data.	77
61	Mosaic of the MTSAT LST product on 15 December 2007 for daytime data.	78
62	Mosaic of the MTSAT LST product on 15 December 2007 for nighttime data.	79
63	Mosaic of the GOES-E LST product on 15 July 2007 for daytime data.	80
64	Mosaic of the GOES-E LST product on 15 July 2007 for nighttime data.	81

List of Tables

1	Summary statistics of the validation results for original AATSR-derived nighttime LST against LST computed from observations at in situ stations.	21
2	Summary statistics of the validation results for original AATSR-derived daytime LST against LST computed from observations at in situ stations.	23
3	Summary statistics of the validation results for AATSR-derived nighttime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.	25
4	Summary statistics of the validation results for corrected AATSR-derived daytime LST against LST computed from observations at in situ stations.	25
5	Summary statistics of the validation results for MODIS-derived nighttime LST against LST computed from observations at in situ stations.	25
6	Summary statistics of the validation results for MODIS-derived daytime LST against LST computed from observations at in situ stations.	26
7	Summary results for the global inter-comparison of AATSR LST derived in WACMOS-ET with the Generalized Split-Window Algorithm with the MODIS MOD11A1 product for one summer and one winter day of 2007	26
8	Summary statistics of the comparison of the WACMOS-ET derived AATSR LST product, calculated for the year 2007 at nine locations with in situ observations.	35
9	Summary statistics of the comparison of the GlobTemperature AATSR LST product, calculated for the year 2007 at nine locations with in situ observations.	36
10	Same as Table 9 but using the same cloud-mask as used for the WACMOS-ET AATSR product.	37
11	Overall summary statistics of all nighttime and daytime matchups for 2007 at all in situ sites. Note that the same cloudmask was used for both products.	37
12	Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for daytime data on 15 July 2007.	46
13	Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for nighttime data on 15 July 2007.	47
14	Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for daytime data on 15 December 2007.	48
15	Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for nighttime data on 15 December 2007.	49
16	Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for daytime data on 15 July 2007	53
17	Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for nighttime data on 15 July 2007	54
18	Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for daytime data on 15 December 2007	56
19	Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for nighttime data on 15 December 2007	57
20	Summary statistics of the validation results for corrected GOES-derived daytime LST against LST computed from observations at in situ stations.	61
21	Summary statistics of the validation results for corrected GOES-derived nighttime LST against LST computed from observations at in situ stations.	62

Summary

The Land Surface Temperature (LST) products generated specifically for the WACMOS-ET project, funded by the European Space Agency (ESA), are evaluated with respect to their overall quality. LST products derived from observations acquired by the Advanced Along-Track Scanning Radiometer (AATSR), the Multi-Functional Transport Satellite (MTSAT), and the Geostationary Operational Environmental Satellite (GOES) were studied here. Following previously established best-practices on LST validation, the evaluation includes both a qualitative component addressing general issues with the data, as well as a quantitative component, which compares the LST products directly against a reference dataset. For the latter, satellite LST is compared against both ground-based in situ datasets acting as a source of absolute reference data and against independent satellite-based LST products from other sensors to provide a spatially exhaustive relative comparison.

In general the results indicate mostly satisfactory performance of the three LST products, although several issues arose during the validation phase. As expected, nighttime LST retrievals far outperform the daytime LST for all instruments.

A direct comparison of AATSR-derived LST against in situ observations indicated a mean nighttime bias of 0.3 °C and a mean daytime bias of 1.4 °C. The standard deviations were found to be 1.3 °C and 2.5 °C, respectively. The root mean squared error (RMSE) as a measure of overall product accuracy was found to be 1.4 °C and 3.2 °C for nighttime and daytime data, respectively. LST derived from AATSR was found to be negatively affected by insufficient cloud masking during nighttime observations. The WACMOS-ET AATSR product was found to provide slightly more accurate retrievals than those of the GlobTemperature AATSR product when the same cloud mask is used.

Due to the lack of usable in situ sites in Australia, MTSAT LST was compared only qualitatively against the available station data and the validation mostly relied on a comprehensive inter-comparison with the MODIS MOD11 LST product. This showed good correspondence for nighttime data. For nighttime data the overall inter-instrument bias was found to be only on the order of 0.1 to 0.2 °C, indicating that MTSAT and MODIS provide essentially the same nighttime LST. For daytime data, however, the biases were found to be between -3.6 °C and -4.0 °C, indicating that the MTSAT LSTs are significantly higher than MODIS LST for daytime retrievals.

LST retrievals from GOES-E showed a good agreement with the ground-based in situ observations. The average bias over 7 stations was found to be 0.37 °C for daytime data and as low as 0.21 °C for nighttime data. The standard deviations were found to be 2.9 °C and 2.4 °C, respectively. The RMSE for daytime data was 3.2 °C and for nighttime data 2.6 °C.

Overall, the three LST products evaluated here show good to acceptable performance within the limits of what is currently achievable with LST products. Nighttime retrievals far outperform daytime retrievals for all instruments so it is recommended to use nighttime data whenever possible.

1 Introduction

Land Surface Temperature (LST) is an important geophysical parameter controlling the fluxes between the Earth's surface and the atmosphere. As such, it is a vital input parameter for estimating evapotranspiration (ET). As part of the WACMOS-ET project, LST was computed using data from several satellite instruments in both geostationary and low-earth orbit. For many applications, but in particular for modelling of evapotranspiration, it is important to have a good overview of the quality of the used LST products.

Here we present the results of the WACMOS-ET LST validation study. The validation was carried out for a total of three instruments, namely the Advanced Along-Track Scanning Radiometer (AATSR), the Multi-Functional Transport Satellite (MTSAT), and the Geostationary Operational Environmental Satellite (GOES) using multiple validation techniques. Firstly, a comparison against ground-based measurements acquired by continuously operating stations provides an absolute reference point of the LST with respect to highly accurate and well-calibrated radiometers. Secondly, an inter-comparison with other LST products was carried out in order to provide a spatial dimension to the validation and to identify spatial patterns in LST accuracy.

In this report, we first briefly discuss the used data and associated methodology in Section 2. The fundamentals of the LST retrieval algorithm are very briefly described in Section 2.1, and the in situ data is described in Section 2.2. Section 2.3 then briefly summarizes the independent remote sensing data used and Section 2.4 describes the processing of the in situ and satellite data. Subsequently, Sections 3 through 5 provide the validation results for AATSR, MTSAT, and GOES, respectively. Finally, Section 7 provides some conclusions and looks at the challenges ahead.

2 Data and Methodology

The overall validation strategy follows the best practices approach described in Schneider et al. (2012). In this document, four main validation techniques are distinguished, namely

- Comparison with in situ measurements (A)
- Radiance-based validation (B)
- Inter-comparison with other products (C)
- Time series analysis (D)

Each category is then further subdivided into a varying number of accuracy classes (Figure 1). Of the four categories listed and described in Schneider et al. (2012), categories A) and C) were applied for validation of the LST products derived within the framework of the WACMOS-ET project. In the following we describe the data sources that were used for generating the respective reference datasets.

2.1 LST retrieval

LST was derived within the framework of the WACMOS-ET project from a total of 4 different instruments, namely from AATSR, MTSAT, GOES-E, as well as the Spinning Enhanced Visible and Infrared Imager (SEVIRI). LST from the first three sensors has been validated as part of WACMOS-ET and the results are being reported on here. LST derived from the SEVIRI

		Category			
		A In situ	B Radiance- based	C Inter-comparison	D Time series
Accuracy Class	Highest accuracy	A1	B1	C1	D1
		A2			D2
		A3	B2	C2	D3
		A4			D4
		A5	B3	C3	
	Lowest accuracy	A6			

Figure 1 – Overview of the various LST validation categories and the corresponding accuracy classes as described in (Schneider et al., 2012). The two LST validation categories that are used primarily in this study are marked in red.

instrument as part of the Land Surface Analysis Satellite Applications Facility (LSA-SAF) is used in WACMOS-ET, however its quality was not specifically evaluated here as it has already been validated comprehensively in several comprehensive studies in the past (Kabsch et al., 2008; Trigo et al., 2008a; Freitas et al., 2010; Göttsche et al., 2013).

The retrieval algorithms used for deriving LST from the satellite-measured brightness temperatures vary to some extent for the various instruments. The algorithm for the instruments with two thermal infrared channels (AATSR, MTSAT, SEVIRI) the retrieval algorithm was based on the Generalized Split-Window approach proposed by Wan and Dozier (1996) and later adapted by Trigo et al. (2008b) and Freitas et al. (2010) but was modified from previous applications by including an additional term to explicitly correct for the effect of the zenith view angle of the various satellite instruments.

As such, the LST or skin temperature T_s is retrieved for AATSR and MTSAT as

$$T_s = C + \left(A_1 + A_2 \frac{1 - \varepsilon}{\varepsilon} + A_3 \frac{\Delta \varepsilon}{\varepsilon^2} \right) \frac{T_{11} + T_{12}}{2} + \left(B_1 + B_2 \frac{1 - \varepsilon}{\varepsilon} + B_3 \frac{\Delta \varepsilon}{\varepsilon^2} \right) \frac{T_{11} - T_{12}}{2} + d(T_{11} - T_{12}) [\sec(\theta) - 1] \quad (1)$$

where C , A_1 , A_2 , A_3 , B_1 , B_2 , B_3 , d are regression coefficients, T_{11} and T_{12} are the top-of-atmosphere brightness temperatures at 10.8 μm and 12.0 μm , θ is the zenith view angle, and ε the average of the respective emissivities ε_{11} and ε_{12} .

In this approach surface temperature is estimated as a linear function of the clear-sky top-of-atmosphere brightness temperatures for the split-window channels at 10.8 μm and 12.0 μm , where the regression coefficients explicitly depend on the land surface emissivity for each channel and further depend implicitly on the atmospheric total column water vapour and zenith view angle. For GOES-12, which includes only a single thermal infrared channel for 10.2-11.2 μm , a new mono-channel algorithm with dynamic emissivity was developed and evaluated. It retrieves LST or skin temperature T_s as

$$T_s = a \frac{T_{TIR}}{\varepsilon} + b \frac{1}{\varepsilon} + c + d [\sec(\theta) - 1] \quad (2)$$

where a , b , c , and d are empirical coefficients that are dependent on the atmospheric water vapour content and zenith view angle, θ is the zenith view angle, T_{TIR} is the brightness temperature in the 10.2-11.2 μm channel, and ε is the surface emissivity in that channel.

Further details about the exact retrieval methodology for each instrument can be found in the WACMOS-ET LST Product Algorithm Theoretical Basis Document (Martins et al., 2014).

2.2 In situ data

Direct validation of satellite-based LST against LST estimates generated from ground-based radiometer observations is the most obvious validation technique. In the following we describe sites which were used for the direct validation against in situ data.

2.2.1 Dedicated in situ validation sites for LST

Gobabeb The Gobabeb site (22.33° S, 15.03° E) is located on large gravel plains (>900 km²) at an altitude of 408 m; these plains are sparsely covered by desiccated grass. To measure LST two self-calibrating KT-15 IR-radiometers are mounted on the 30m tower taking measurements in the 9.6 μm to 11.5 μm range with a target accuracy of ± 0.3 K (Olesen and Göttsche, 2009). The two radiometers observe neighbouring areas of the gravel surface from the north with a 30° viewing angle. Such a viewing angle is justified since angular anisotropic emissivity values for sand, grass, and gravel are negligible up to 30°. An additional KT-15 radiometer measures the effective brightness temperature of the sky at 53° to zenith. Figure 2 shows a photograph of the Gobabeb LST validation site. The site has been providing continuous measurements of LST since January 2008, which is unfortunately just after the end of the WACMOS-ET study period. However, some non-continuous measurements were already taken during December 2007 and these were used for validation of AATSR at Gobabeb. Unfortunately only a very small number of matchups were available. Further details about the site can be found in Göttsche et al. (2011, 2013).

Evora The Evora site (38.54° N, 8.00° W) is part of the global flux network of measurements (Fluxnet) as is composed of sparse oak tree canopy (30-40 trees/ha) and a grassland soil (Guillevic et al., 2013). In addition to instruments taking meteorological measurements three infrared radiometers (Heitronics KT-15.85 IIP) measure observe sunlit background, tree crown, and the sky at 53° zenith angle from two 10 m high towers. Brightness temperatures of these endmembers are taken in the 9.6 μm to 11.5 μm spectral range, with a target accuracy of 0.3 K (Göttsche et al., 2013; Ermida et al., 2014). For the comparison with satellite-derived LST the in situ brightness temperatures were calculated as a weighted average of the brightness temperatures of the two endmembers soil/grass and tree crown, with the former weighted at 0.68 and the latter weighted at 0.32. The Evora station has been providing LST data continuously since January 2009, however some data was also available during the WACMOS-ET study period for November and December 2007.

Lake Tahoe Unfortunately, the AATSR retrieval algorithm used for WACMOS-ET excluded pixels over inland water bodies and therefore the Lake Tahoe site could not be used for AATSR validation. It was originally planned to use the dataset for validation of LST from GOES, but the site was too far west to be usable for validation with GOES-E. Nonetheless, the description of the site is kept in here as another example of a dedicated LST validation site.

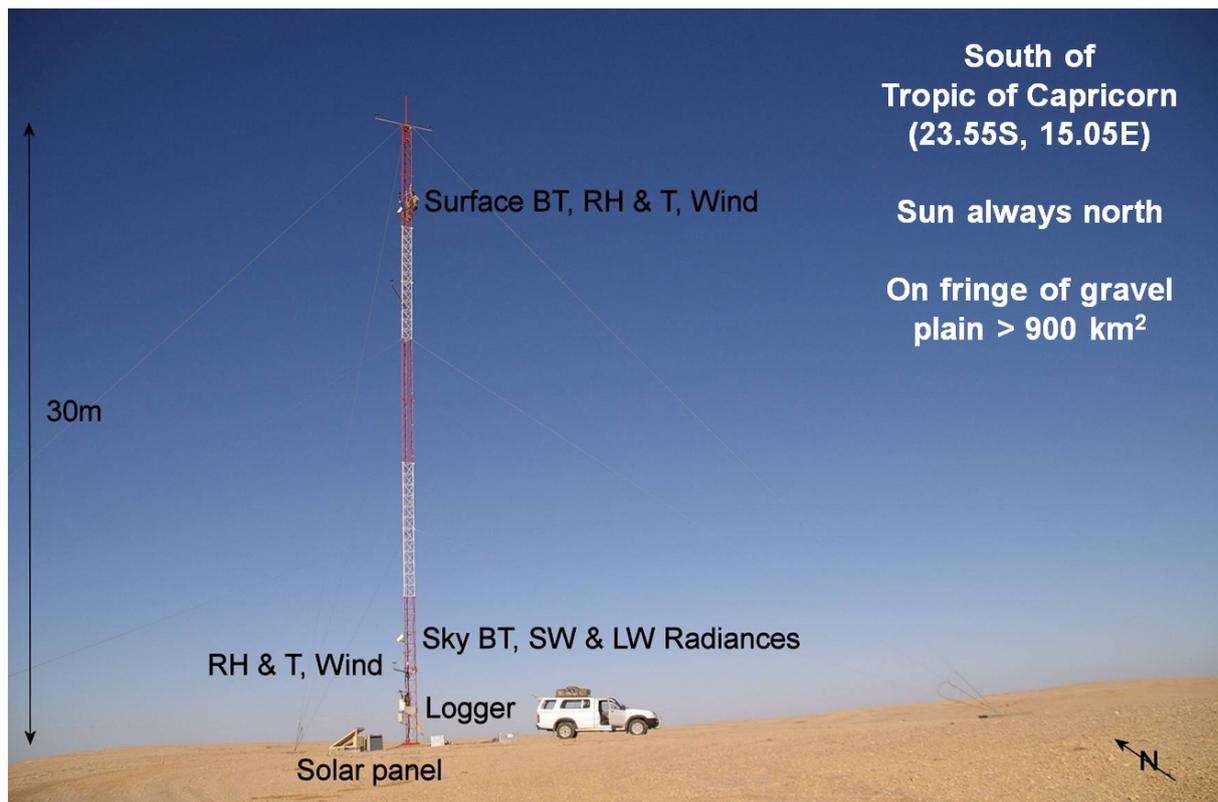


Figure 2 – The Gobabeb LST validation site operated by the Karlsruhe Institute of Technology (Göttsche et al., 2013)

The Lake Tahoe, California/Nevada, automated validation site, which is operated by the Jet Propulsion Laboratory (Hook et al., 2003, 2007), is a dedicated site for LST validation. Due to the extremely homogeneous surface, the accurately known emissivity, and the long-term continuous measurements at a high temporal sampling frequency, the site provides an ideal-case scenario for LST validation.

Four buoys (TB1 through TB4) located at Lake Tahoe have been measuring LST and a wide variety of meteorological and limnological parameters at an interval of 2 minutes since the year 1999. Figure 3 shows the locations of the four buoys within the lake. Each buoy provides radiometer measurements observed using a custom-built radiometer with an accuracy of ± 0.1 °C. Figure 4 shows a single instrument platform mounted on one buoy. The radiometer observations were subsequently corrected for the emissivity of water and the downwelling sky radiance reflected by the surface. The latter was accomplished by using a radiative transfer model driven by atmospheric profiles obtained from the NCEP reanalysis (Kalnay et al., 1996). Details on the station and the processing methodology can be found in Hook et al. (2003).

2.2.2 SURFRAD sites

Due to the scarcity of dedicated in situ LST sites in general, and in particular during the study period, the validation of the LST component had to rely to some extent on data provided by non-dedicated station that use radiometers to accurately measure upwelling and downwelling longwave thermal infrared radiation.

Such stations exist for example within the Surface Radiation (SURFRAD) network (Augustine et al., 2000, 2005), which is a collection of stations scattered throughout the United States.

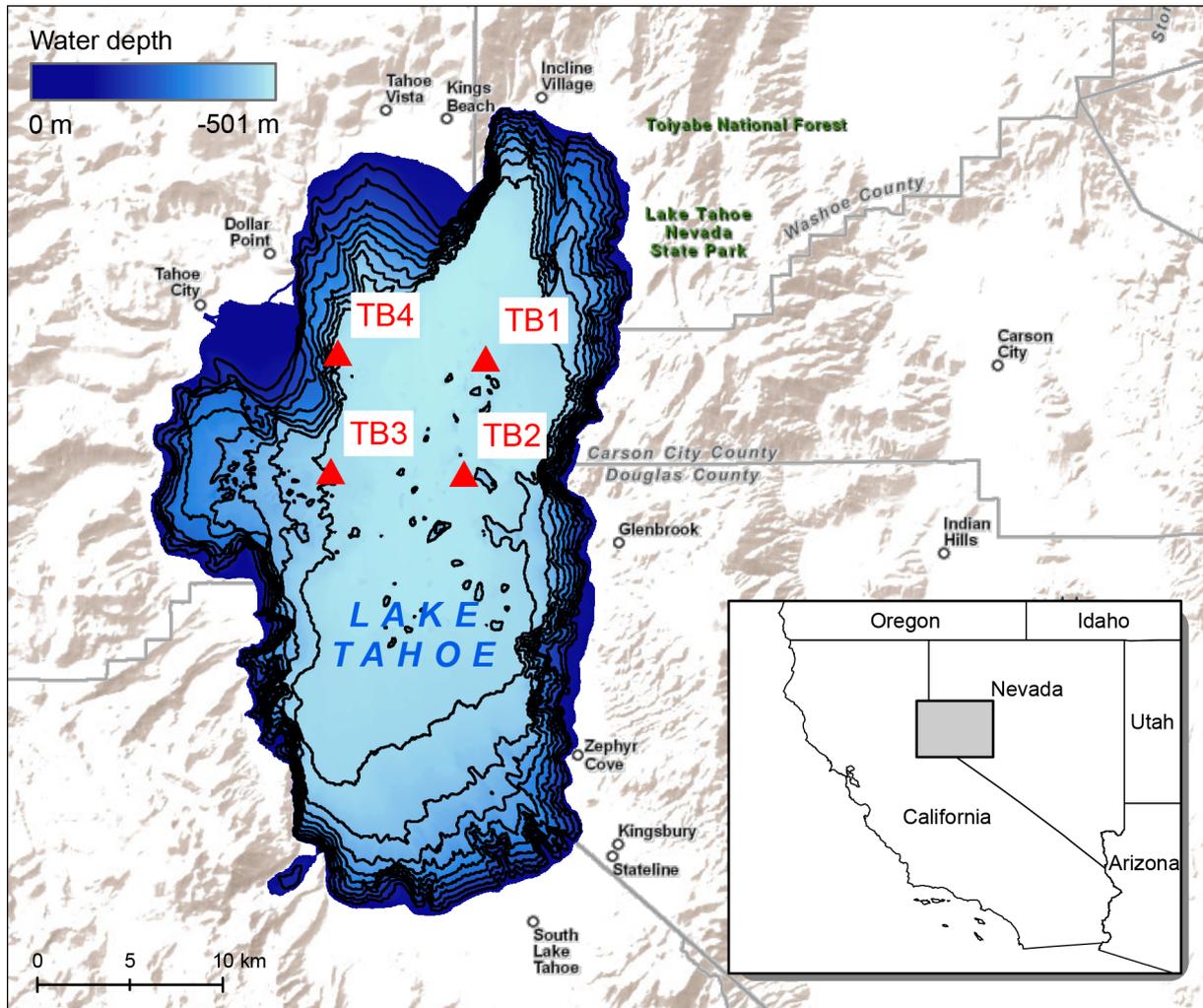


Figure 3 – Geographic overview and bathymetry of the Lake Tahoe Study Site. Contour lines are at 100 m intervals. The red triangles labeled TB1 through TB4 indicate the location of the four buoys operated by the Jet Propulsion Laboratory. Background map provided by Environmental Systems Research Institute. Bathymetry data provided by United States Geological Survey.



Figure 4 – One of the four Lake Tahoe buoys measuring LST as well as a variety of meteorological parameters.

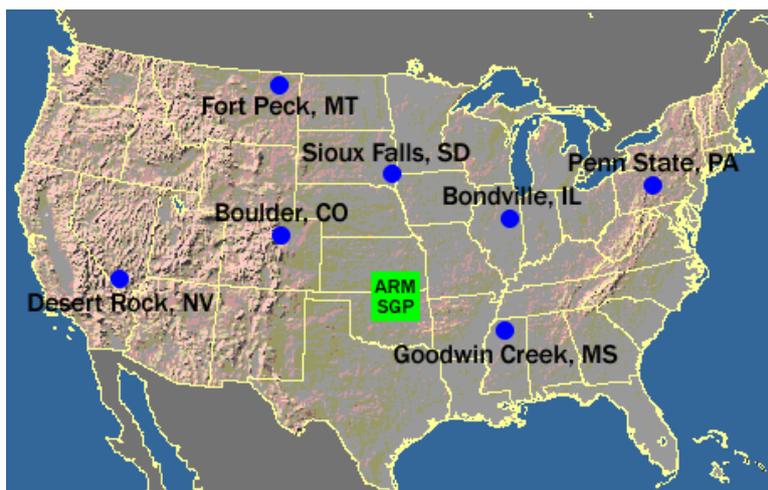


Figure 5 – Map of the location of the SURFRAD sites within the contiguous United States. (from <http://www.esrl.noaa.gov/gmd/grad/surfrad/sitepage.html>)

Figure 5 shows the spatial distribution of the currently existing stations. Currently seven stations in various U.S. states are operational. The primary objective of the SURFRAD sites is to support climate research with accurate, continuous, and long-term measurements. Figure 6 shows an example of a SURFRAD station (Desert Rock in Nevada).

It should be noted that the locations of the SURFRAD stations were not chosen with validation of satellite-based LST products in mind and therefore are not in all cases as homogeneous as dedicated LST validation stations such as Gobabeb or Lake Tahoe. As such the uncertainty in estimated emissivity will always be higher at the SURFRAD stations. Nonetheless the sites have been used successfully for LST validation in the past (Wang and Liang, 2009; Guillevic et al., 2012, 2014).

2.2.3 ARM sites

In addition to SURFRAD stations, sites organized within the ARM Climate Research Facility (Stokes and Schwartz, 1994) have been used. This network operates several field research sites worldwide to study the effects of aerosols, precipitation, surface flux, and clouds on global climate change. Most sites are equipped with a range of instrumentation including upward and downward-looking ground radiometers for measuring both the brightness temperatures (BTs) of the sky and surface respectively. In addition, the sites also house a range of standard meteorological equipment, such as humidity sensors, thermometers, and ceilometers. The downward-looking radiometers take average measurements every 60 seconds of the surface.

A feature of the ARM sites is that in general the data sets recorded are near continuous. The measurement period varies from site to site, however most sites have been providing data for at least several years. It should be noted that, while the ARM sites provide a reasonable reference dataset, they are not dedicated LST validation sites and as such do not in all cases exhibit homogeneous land cover at the scale of the satellite footprint.

2.3 Independent remote sensing data

In addition to absolute comparisons against ground-based reference data from point sources, validation of satellite LST was performed against LST datasets derived from other spatially



Figure 6 – Example of a SURFRAD station, here showing the Desert Rock (Nevada) station. Clockwise from left center: Radiometer platform, met tower, total sky imager, and solar tracker. From <http://www.esrl.noaa.gov/gmd/grad/surfrad/>.

distributed data sources. This type of validation allows for the detection of spatial patterns in the quality of the LST data and is a valuable complement to in situ-based validation, which is relatively limited in its spatial scope. It further allows the detection and monitoring of inter-sensor biases, which is essential when merging data from multiple satellite instruments.

In the context of WACMOS-ET, data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) was used for this purpose. More specifically, the MOD11A1 product was used for inter-comparison with the products generated for WACMOS-ET. This product (Wan, 2007) is based on the generalized split-window algorithm (Wan and Dozier, 1996) and has been validated extensively (Wan et al., 2002, 2004; Wan and Li, 2008; Wan, 2008; Coll et al., 2005, 2009; Galve et al., 2007). The results generally indicate errors of less than 1 K on homogeneous surfaces. Validation of the MOD11 product has further been carried out using the same Lake Tahoe in situ dataset described in the previous section (Schneider et al., 2009; Schneider and Hook, 2010).

The MOD11A1 products is delivered in tiles projected in the sinusoidal projection. This is identical to the projection and tiling approach used in WACMOS-ET. Figure 7 shows an example of the type of MOD11A1 tiles which were used to compare the WACMOS-ET LST data against.

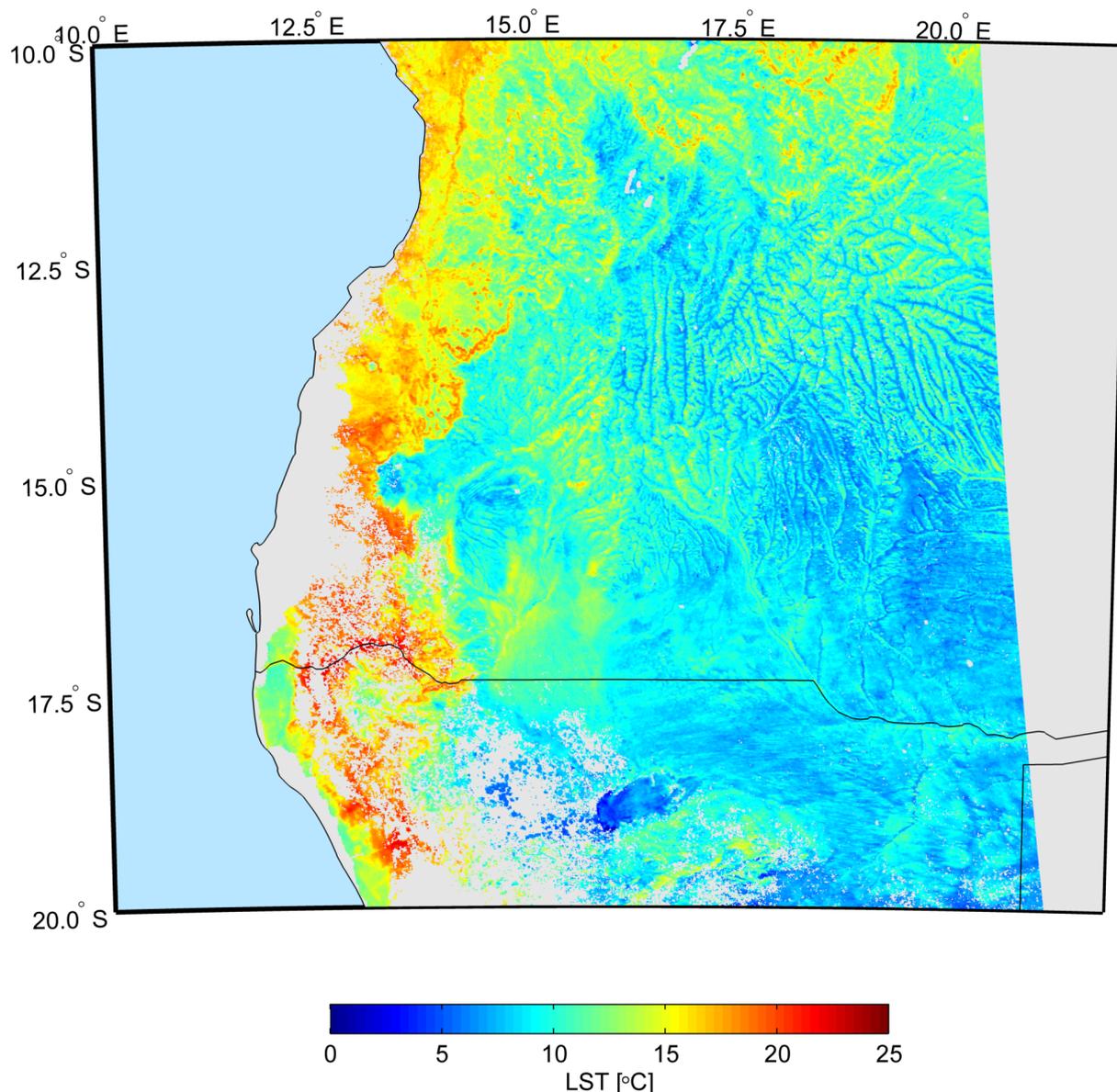


Figure 7 – Example of one tile of the MOD11A1 LST product, here showing the nighttime LST data for the tile h19v10 along the border between Namibia and Angola. Gray areas indicate land surfaces over which LST could not be retrieved due to clouds or the retrieval was of low quality.

In addition to MODIS data, the AATSR LST product developed as part of the ESA-funded GlobTemperature project (www.globtemperature.info) has been used to evaluate the AATSR LST product used within the framework of WACMOS-ET. The GlobTemperature product is based on an algorithm developed by Prata (2002) but has been improved with respect to multiple aspect including the spatial resolution of the auxiliary datasets such as fractional vegetation cover and global biome distribution. Only GlobTemperature data for the year 2007 was evaluated in this study.

2.4 Processing

This section briefly summarizes the processing steps that needed to be carried out for the in situ LST data as well as the independent remote sensing data. While not a large amount of processing was required for the preparation of the independent remote sensing data, considerable effort

was in contrast necessary to produce realistic in situ LST estimates from those in situ stations which only provided upwelling and downwelling longwave thermal infrared radiation.

2.4.1 Processing of in situ data

Only for the Gobabeb and Evora stations in situ LST was already pre-computed using local expert knowledge and locally measured emissivities. When in situ LST was not already available, such as for example for the SURFRAD sites, it was computed from the radiometer-based brightness temperatures, the sky radiance that is reflected by the Earth surface into the viewing geometry of the radiometer, and the thermal emissivity of the surface. The upwelling spectral radiance measured by the radiometer is defined as

$$LW_{out} = (1 - \varepsilon)LW_{in} + \varepsilon\sigma T_s^4 \quad (3)$$

where LW_{out} and LW_{in} are the upwelling and downwelling longwave radiation, respectively, ε is the broadband emissivity of the Earth's surface in the spectral range of measurements, σ is the Stefan Boltzmann constant ($5.6704 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$), and T_s is the surface temperature.

Based on Equation 3, the in situ skin land surface temperature T_s is then calculated as

$$T_s = \left[\frac{LW_{out} - (1 - \varepsilon) \cdot LW_{in}}{\varepsilon \cdot \sigma} \right]^{\frac{1}{4}}. \quad (4)$$

The emissivity at the various stations (with exception of Gobabeb and Evora, for which it was already available through comprehensive measurements by the data providers) was obtained from the Global Infrared Land Surface Emissivity Database (Seemann et al., 2008) operated by the Cooperative Institute for Meteorological Satellite Studies (CIMSS) and available at <http://cimss.ssec.wisc.edu/iremisis/>. It provides MODIS-derived global estimates of land surface emissivity at a 0.05 degree spatial resolution. The data is supplied for each month from 2003 to present, and as such is able to account for seasonal variability in the emissivity value at each station.

Broadband emissivity was estimated from the CIMSS dataset following the approach suggested by Ogawa et al. (2003) and Wang et al. (2005) as a linear combination of three narrowband emissivities as

$$\overline{\varepsilon}_w = a \cdot \varepsilon_{8.5} + b \cdot \varepsilon_{11} + c \cdot \varepsilon_{12} \quad (5)$$

where $\overline{\varepsilon}_w$ is the broadband emissivity, $\varepsilon_{8.5}$, ε_{11} and ε_{12} are the narrowband emissivities at 8.5 μm , 11 μm , and 12 μm , respectively. The coefficients a , b , and c were found to be 0.2122, 0.3859, and 0.4029 respectively (Wang et al., 2005).

The sub-pixel variability of emissivity and thus LST is a concern when comparing satellite-derived LST, which is given for an area on the order of one to several square kilometers, with radiometer-based in situ data which are essentially point measurements and at most cover an area of a few square meters. Aside from special-purpose sites dedicated to LST validation such as Gobabeb or Lake Tahoe, nearly all sites exhibit a certain amount of spatial heterogeneity in emissivity. Current work is investigating possible strategies on how to overcome this issue (Guillevic et al., 2012), however the required methodologies exceeded the scope of this validation exercise and were therefore not considered as part of this work. Validating satellite-derived LST against non-dedicated in situ sites can still be helpful in determining the real-world uncertainty in the retrievals, as most applications of LST products take place in areas of highly variable land cover and thus emissivity. LST validation is further complicated by variable viewing angles which have an impact on the errors and the inter-comparison between sensors. This is further exacerbated by complex topography which can cause shadowing effects at low view angles.

2.4.2 Processing of independent remote sensing data

Not much processing had to be carried out on the independent remote sensing data. The MODIS-Terra based MOD11A1 product came in HDF4 format and was analyzed and compared with the LST data from the instruments in the same sinusoidal projection and the same tile-system that the product came in. Following best practice guidelines outlined in Schneider et al. (2012), the inter-comparison with independent remote sensing data was carried out by resampling the higher-resolution dataset (e.g. AATSR) to the same spatial grid as the data from the lower-resolution instrument (e.g.. MTSAT and GOES).

3 Validation of AATSR

3.1 Validation with in situ observations

Validation using in situ observations of LST were carried in order to evaluate the performance of the WACMOS-ET LST product with respect to an absolute reference dataset. It should be noted that with the exception of the Gobabeb station operated by KIT and the Lake Tahoe station operated by JPL, none of the available stations are dedicated LST validation stations. As such, they are not ideal for this purpose and are subject to several drawbacks, with the primary issue being that the surroundings of the stations are generally not homogeneous at the scale of satellite pixels and thus the assumption made for the emissivity can only be a very approximate guess. In addition the emissivity derived from the CIMSS dataset (Seemann et al., 2008) is also subject to uncertainty.

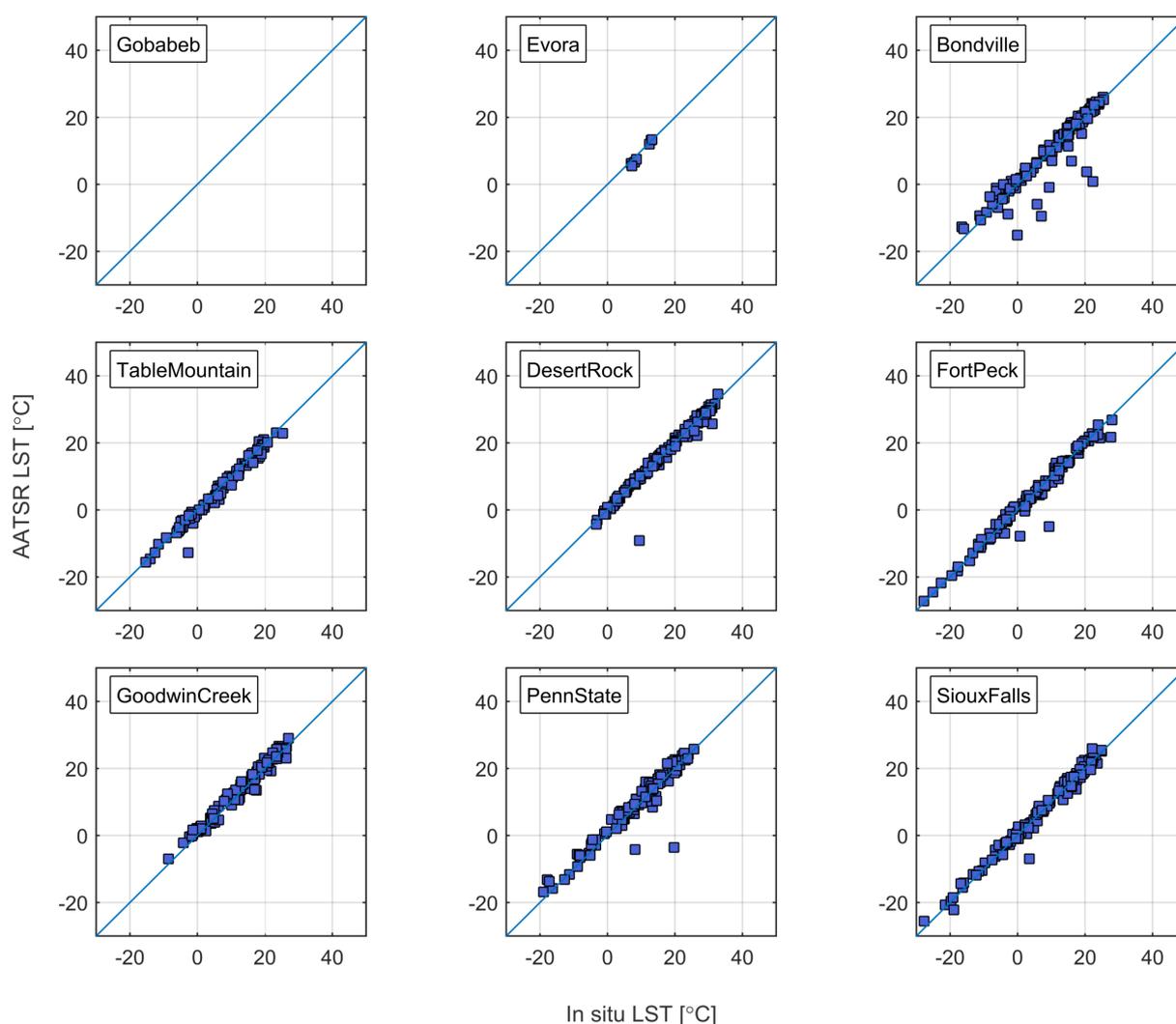


Figure 8 – Scatter plots of original AATSR-derived nighttime LST against LST computed from observations at in situ stations.

Figure 8 shows scatterplots of the in situ LST versus AATSR LST for nighttime matchups. It can be observed that the majority of matchups follows the 1:1 line at all sites, but also that at most sites frequent negative outliers occur. These are the characteristics of failures of the used cloud mask. Another indicator for this behavior is that a site located in the Nevada desert (Desert

Table 1 – Summary statistics of the validation results for original AATSR-derived nighttime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std. Dev.	RMSE
Gobabeb	0			
Evora	7	-0.7	0.8	1.0
Bondville	115	-0.1	4.1	4.1
TableMountain	92	-0.6	1.5	1.6
DesertRock	137	-0.1	1.9	1.9
FortPeck	114	-0.1	2.0	2.0
GoodwinCreek	107	0.9	1.4	1.7
PennState	103	0.7	3.2	3.3
SiouxFalls	113	0.3	1.7	1.7
Average	87.6	0.0	2.1	2.1

Rock) which sees relatively few clouds has a much smaller number of negative outliers than the typically more cloudy sites in the mid-latitudes (e.g. the Bondville site).

Clearly the cloud masking procedure applied here, which makes use of the cloud flags in the original L1 AATSR data, is not ideal for nighttime LST retrievals. However, it was outside of the scope of the project to develop more reliable cloud masks. A lot of work has been done along those lines in the past (e.g. Merchant et al., 2005; Bulgin et al., 2014) and similar methods could be used to improve the cloud masking. Doing so would significantly improve the accuracy of the LST product.

The quantitative results shown in Table 1 reiterate these issues with the cloud masking. For most stations, the bias is slightly negative, indicating that the estimated AATSR LST is less than the reference LST observed at the respective stations. The lowest absolute bias was found for the Desert Rock and Fort Peck stations. The highest absolute bias for a station with sufficient number of samples was observed for the Goodwin Creek station.

Compared to its scatterplot in Figure 8, the Bondville station has a surprisingly low bias of only -0.1 °C. This is initially surprising as the scatterplot clearly shows a large number of negative outliers. However, the relatively large number of matchups which have been properly cloudmasked are biased slightly high at this station and thus to some extent compensate for the few highly negative outliers due to faulty cloud masking issues. What this means in practice, however, is that the true bias of the properly cloud-cleared matchups cannot be determined accurately as it is masked by the negative outliers due to issues with an insufficient cloudmask, particularly for nighttime data.

Figure 9 shows similar scatterplots but for daytime matchups only. The situation for daytime LST is quite different. While a few isolated cloud masking issues can be seen here as well, particularly for the Fort Peck station and to some extent at the Desert Rock station, they are substantially reduced in number in comparison to the nighttime retrievals. This is of course due to the additional availability of the visible bands during daytime retrievals which significantly simplify the cloud masking problem. The scatter follows the 1:1 line for most stations but a clear tendency towards higher errors for higher temperatures greater than 30 °C is visible, particularly at the Bondville and Sioux Falls stations. At some stations, such as Table Mountain and Fort Peck the AATSR LSTs also shows a tendency of overestimating the in situ LST for temperatures greater than 20 °C whereas the bias is much closer to zero for temperatures less than that value.

The quantitative evaluation for the daytime data (Table 2) reflects the existence of these systematic outliers. At most stations except Goodwin Creek and Evora, which has only a very small number of matchups and is only suitable for validation of daytime LST when the view-angle

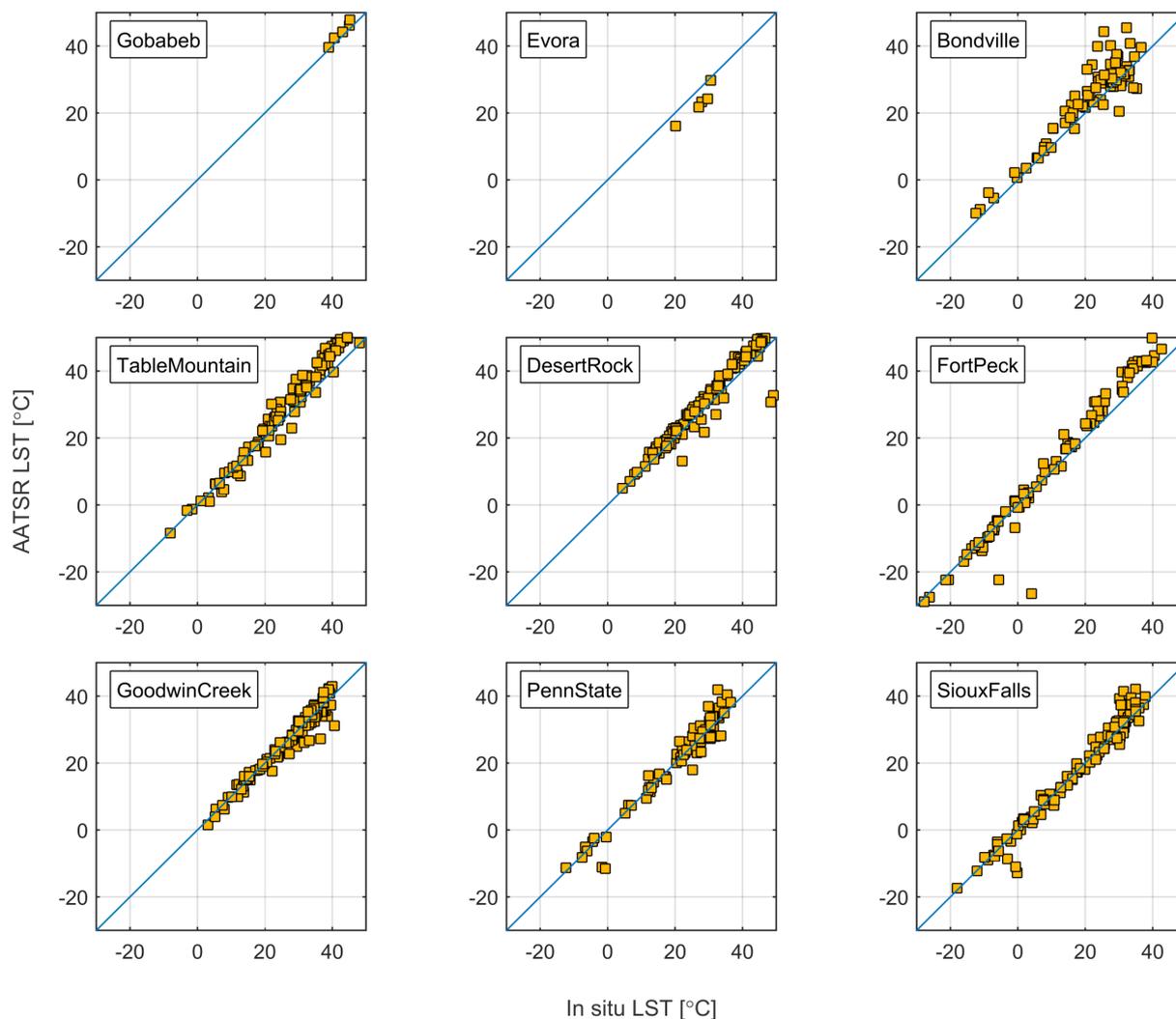


Figure 9 – Scatter plots of original AATSR-derived daytime LST against LST computed from observations at in situ stations.

dependent variable endmember fractions are taken into account appropriately (Ermida et al., 2014; Guillevic et al., 2013), the bias is positive, indicating that AATSR LST slightly overestimates the LST based on the in situ observations. The bias is highest (>3 °C) at stations which show the aforementioned systematic errors (primarily Table Mountain and Bondville), whereas other stations, for example Goodwin Creek, Penn State, and Sioux Falls reach more acceptable biases of between 0 °C and 1 °C.

Because the cloud masking issue resulted in a unusually high number of outliers which in turn masked the true accuracy of the AATSR-based LST retrievals, a very simple filtering schemes was used to eliminate at least the most extreme outliers caused by an inefficient cloud masking procedure. This scheme evaluated the difference between in situ LST and AATSR LST and removed those matchups where this difference was at the same time negative and its value greater than two standard deviations from the mean. While this is a relatively crude method it is effective here in the sense that it allows for the elimination of gross outliers, which are primarily caused by cloud masking issues. Thus it becomes possible to derive summary statistics which are more representative of the actual accuracy of the AATSR LST product once the nighttime cloud masking issues are taken care of.

Table 2 – Summary statistics of the validation results for original AATSR-derived daytime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Gobabeb	5	1.5	0.8	1.6
Evora	5	-4.0	1.9	4.4
Bondville	79	3.2	4.7	5.6
TableMountain	111	3.1	3.3	4.5
DesertRock	134	2.3	3.5	4.2
FortPeck	81	2.0	5.4	5.7
GoodwinCreek	102	-0.3	2.3	2.3
PennState	74	0.4	3.2	3.2
SiouxFalls	105	0.4	3.0	3.0
Average	77.3	1.0	3.1	3.8

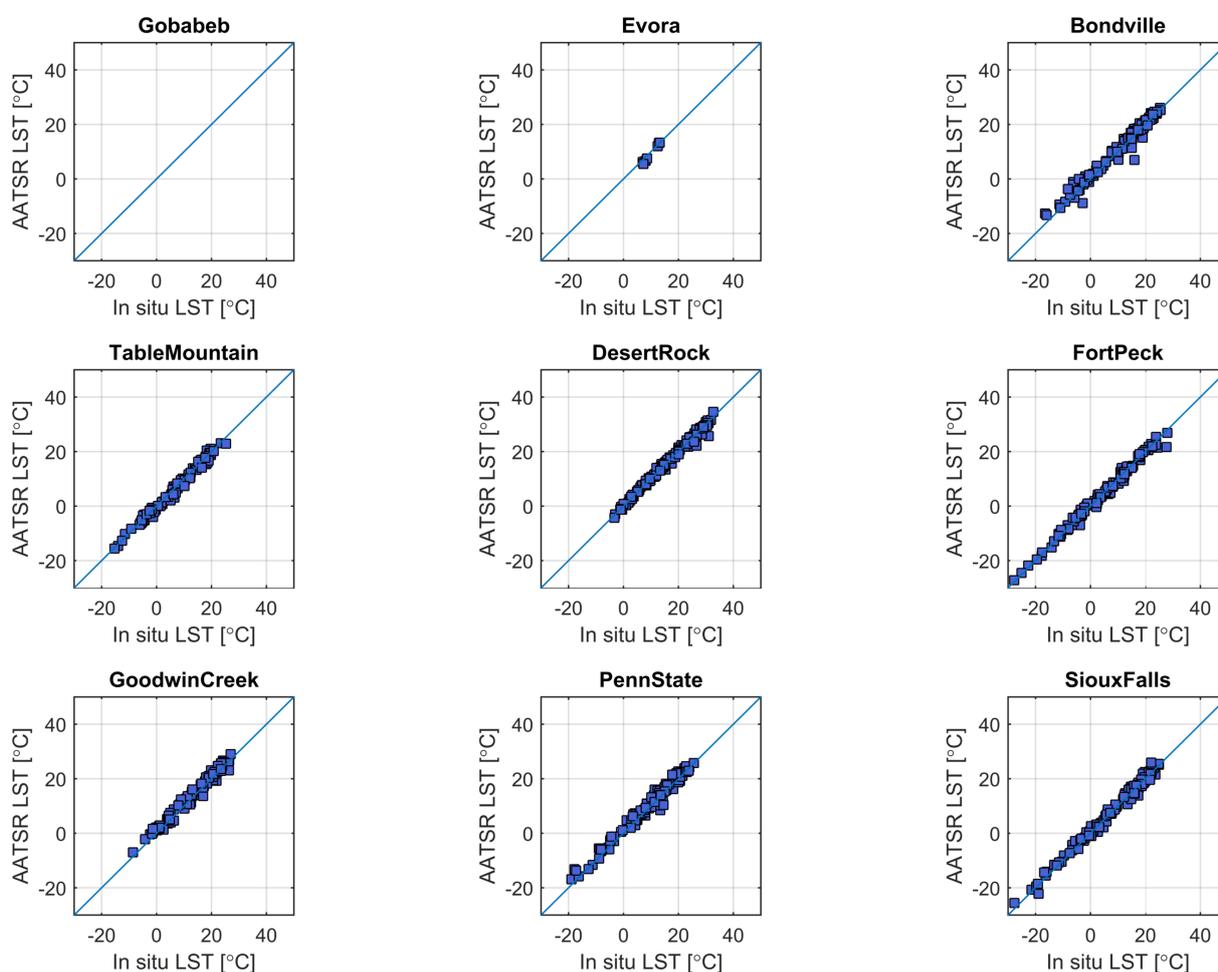


Figure 10 – Scatter plots of AATSR-derived nighttime LST against LST computed from observations at in situ stations after additional statistical cloud filtering. Compare to unfiltered data in Figure 8.

Figures 10 and 11 show the scatterplots for all stations after the filtering was carried out for nighttime and daytime matchups, respectively. While only a few data points have been removed for for the daytime data, quite a few outliers were removed for the nighttime data.

The quantitative impact of the filtering becomes clear in Tables 3 and 4 which show the respective nighttime and daytime statistics for the various stations. It is quite obvious from the direct comparison between Tables 1 and 3 that the nighttime AATSR retrievals profit significantly

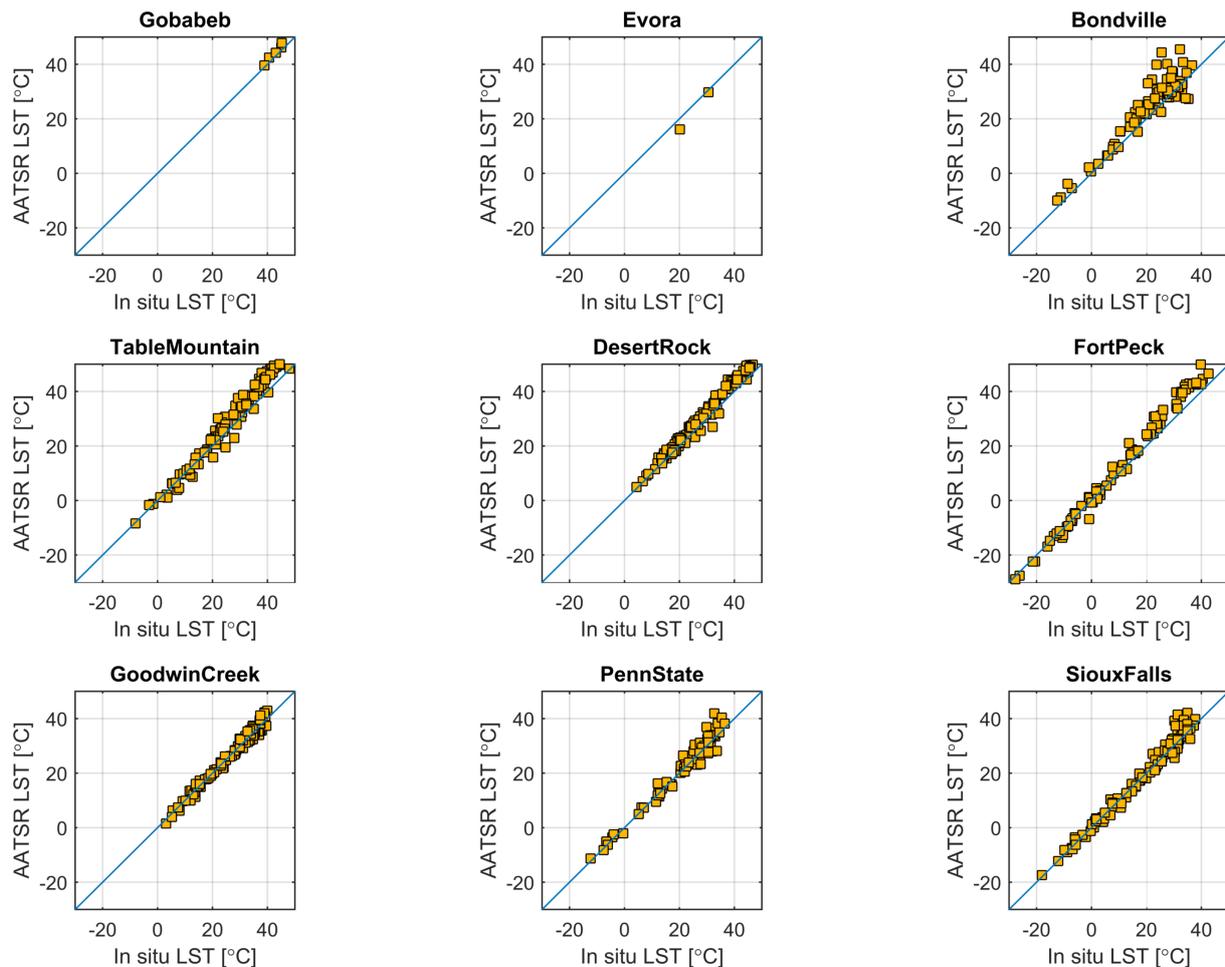


Figure 11 – Scatter plots of AATSR-derived daytime LST against LST computed from observations at in situ stations after additional statistical cloud filtering. Compare to unfiltered data in Figure 9

from the outlier removal. While the average bias has slightly increased, the random error as expressed by the standard deviation has dropped dramatically from 2.1 °C to 1.3 °C. Similarly the RMSE as a measure of overall accuracy has been reduced from 2.1 °C to 1.4 °C.

Since the daytime AATSR LST was much less affected by outliers related to the cloud masking procedure, the filtering did not change as much as for the nighttime data. A visual comparison between Figures 9 and 11 shows that only a few data points were removed. Nonetheless, a comparison of Figures 2 and 4 shows that the statistics have improved. While the average bias has increased slightly from 1.0 °C to 1.4 °C, the average random error has decreased significantly from 3.1 °C to 2.5 °C. Accordingly, the RMSE has decreased as well from 3.8 °C to 3.2 °C.

For comparison and reference, Tables 5 and 6 show corresponding results for the MODIS MOD11A1 product. They show that MODIS-Terra tends to underestimate both the nighttime and daytime LST with average biases of -3.0 °C and -1.5 °C, respectively. Such negative biases were reported previously for MODIS-Terra by (Wang et al., 2008), although they later on found a reduction in bias for MODIS-Aqua (Wang and Liang, 2009). Further investigation will be necessary to determine if there is a significant difference in the accuracy between the LST products from Terra and Aqua.

Table 3 – Summary statistics of the validation results for AATSR-derived nighttime LST against LST computed from observations at in situ stations after additional statistical cloud filtering. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Gobabeb	0			
Evora	7	-0.7	0.8	1.0
Bondville	109	0.7	1.9	2.0
TableMountain	91	-0.5	1.1	1.2
DesertRock	136	0.0	1.0	1.0
FortPeck	112	0.1	1.3	1.2
GoodwinCreek	106	1.0	1.3	1.6
PennState	101	1.0	1.7	2.0
SiouxFalls	112	0.4	1.3	1.4
Average	86.0	0.3	1.3	1.4

Table 4 – Summary statistics of the validation results for corrected AATSR-derived daytime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Gobabeb	5	1.5	0.8	1.6
Evora	2	-2.5	2.3	2.9
Bondville	78	3.3	4.5	5.5
TableMountain	111	3.1	3.3	4.5
DesertRock	130	2.8	2.1	3.5
FortPeck	79	2.6	3.4	4.3
GoodwinCreek	94	0.2	1.5	1.6
PennState	71	0.8	2.6	2.7
SiouxFalls	102	0.7	2.4	2.5
Average	74.7	1.4	2.5	3.2

3.2 Validation against independent EO data

In the following section the WACMOS-ET derived AATSR LST product is compared against the MODIS MOD11 product as well as against the AATSR LST product generated within the framework of the GlobTemperature project.

Table 5 – Summary statistics of the validation results for MODIS-derived nighttime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Gobabeb	20	-3.3	0.5	3.3
Evora	27	-1.1	1.0	1.4
Bondville	303	-2.7	4.5	5.2
Table Mountain	358	-3.5	3.6	5.0
Desert Rock	39	-4.1	3.2	5.2
Fort Peck	401	-3.2	4.5	5.5
Sioux Falls	373	-3.2	4.5	5.5
Average	217.3	-3.0	3.1	4.5

Table 6 – Summary statistics of the validation results for MODIS-derived daytime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Gobabeb	23	-3.1	1.4	3.3
Evora	29	-5.6	1.5	5.8
Bondville	316	0.2	4.3	4.3
Table Mountain	361	-0.2	2.4	2.4
Desert Rock	7	-0.8	1.7	1.8
Fort Peck	295	1.1	2.5	2.8
Sioux Falls	330	-1.9	2.6	3.2
Average	194.4	-1.5	2.3	3.4

3.2.1 Comparison with MODIS MOD11

The AATSR-based LST product derived in the framework of the WACMOS-ET project was inter-compared with MODIS MOD11A1 data. Table 7 shows the inter-comparison results for the AATSR LST based on the generalized split-window approach. They provide statistics of accuracy (bias), precision (standard deviation) as well as overall performance expressed as the root mean squared error (RMSE).

The results indicate that the the AATSR LST provides consistently higher values than the MODIS MOD11 LST, ranging from 0.4 K to 2.1 K. This would be expected after the comparison against the in situ observations, which indicated a negative bias of the MOD11 product. The RMSEs between MODIS and AATSR LST are consistently below 3 °C.

Table 7 – Summary results for the global inter-comparison of AATSR LST derived in WACMOS-ET with the Generalized Split-Window Algorithm with the MODIS MOD11A1 product for one summer and one winter day of 2007. Bias is computed as MODIS LST minus AATSR LST, i.e. a negative bias indicates that AATSR LST is higher than MODIS LST. “Good tiles” are defined as tiles with a minimum of 1000 valid difference pixels between MODIS and AATSR.

			Bias [K]	Std. Dev. [K]	RMSE [K]
15-Jul-07	Day	All Tiles	-1.02	1.98	2.93
		Good Tiles	-1.03	2.22	2.99
	Night	All Tiles	-2.03	1.52	2.78
		Good Tiles	-2.13	1.66	2.94
15-Dec-07	Day	All Tiles	-0.61	1.90	2.51
		Good Tiles	-0.43	1.92	2.31
	Night	All Tiles	-1.56	1.55	2.73
		Good Tiles	-1.56	1.57	2.50

As an example illustrating the capability of the AATSR and MODIS LST products over time, Figures 12 and 13 shows the full time series of nighttime and daytime LST of the two products against the reference in situ LST for the Bondville and Table Mountain stations, respectively. In general the two LST products are able to roughly delineate the lower (nighttime) and upper (daytime) boundaries of the temperature range measured by the in situ station. A visual inspection indicates that the two products are roughly comparable but that the AATSR product shows slightly higher temperature values than the MODIS product, in particularly for the nighttime data. This is consistent with the results found in the global inter-comparison. As would be expected the scatter for nighttime data is generally lower than for daytime data due to the lack of diurnal heating and rapid temperature changes due to intermittent cloud cover etc. This is

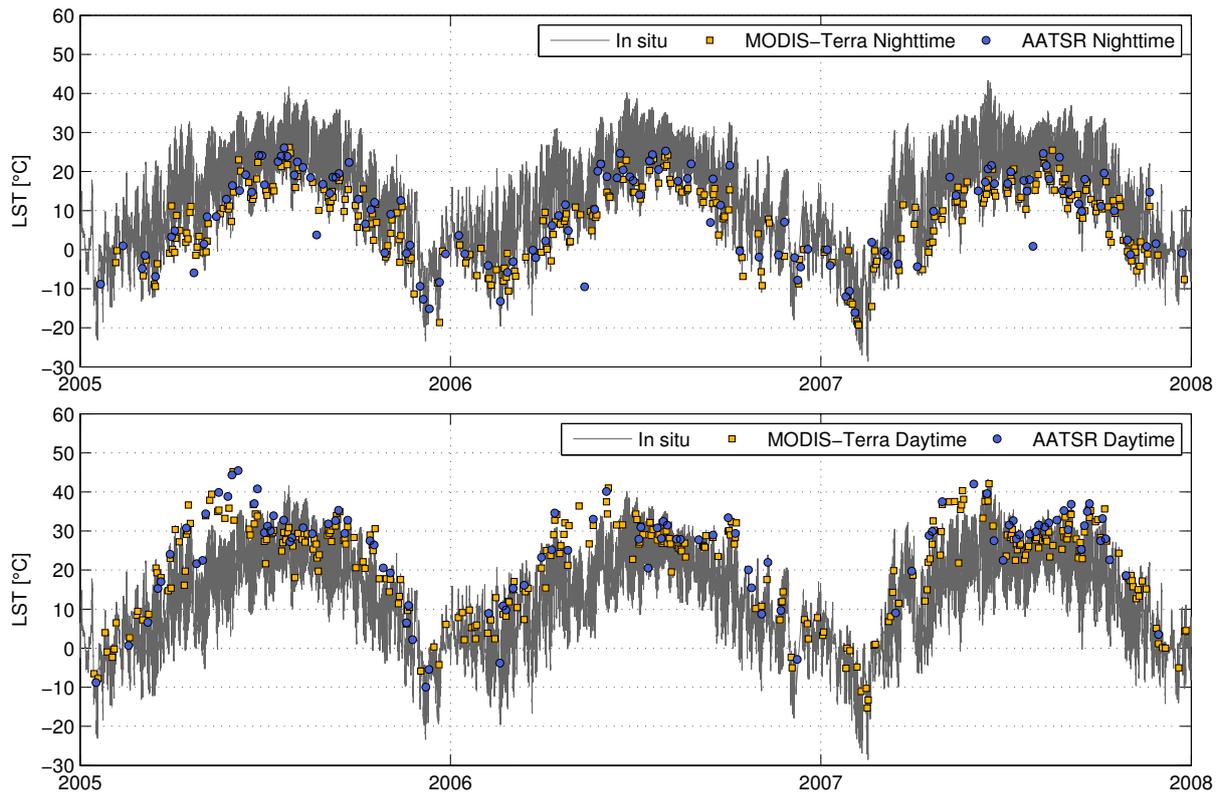


Figure 12 – Inter-comparison of the full nighttime (top) and daytime (bottom) LST time series derived from AATSR versus MODIS-Terra for the entire WACMOS-ET study period at the Bondville station. For reference the continuous in situ LST time series measured at the Bondville station is also given.

particularly obvious for the data at the Table Mountain station (Figure 13). Once again, the cloud masking issues of the AATSR nighttime product are obvious. Both locations show several gross negative outliers in the nighttime LST which are clearly caused by clouds that remained undetected by the cloud masking algorithm. A visual inspection indicates about 4 nighttime outliers for the Bondville station and 1 nighttime outlier for the Table Mountain station. Interesting is also that the satellite-based daytime LSTs at the Bondville station show positive deviation with respect to the in situ data during the spring months. This is true for both the AATSR and the MODIS product so is likely not to indicate an error in the LST retrieval but rather an inconsistency in what the satellite and the in situ station measure during the period, e.g. it could mean that the average emissivity at the spatial scale of the satellites is not comparable to the emissivity at the station during these months due to annual changes in land cover/phenology.

At the Bondville station the median nighttime LST over the entire WACMOS-ET study period was 9.75 °C and 11.39 °C for MODIS and AATSR respectively. For the daytime LST the corresponding median values were 26.28 °C and 28.55 °C. Interestingly, at the Table Mountain station, the median nighttime LST derived from MODIS was 8.16 °C and thus slightly higher than the AATSR-derived median LST of 7.16 °C. The daytime LST at this location was remarkably similar for both satellite products with 31.6 °C for MODIS and 31.3 °C for AATSR. It should be noted that in this comparison not the same number of observations points were considered for both satellite products.

AATSR LST derived within the framework of the WACMOS-ET project was further compared against MODIS data by means of global difference images in order to highlight the spatial patterns in the discrepancies between the retrieved LST from the two data sources.

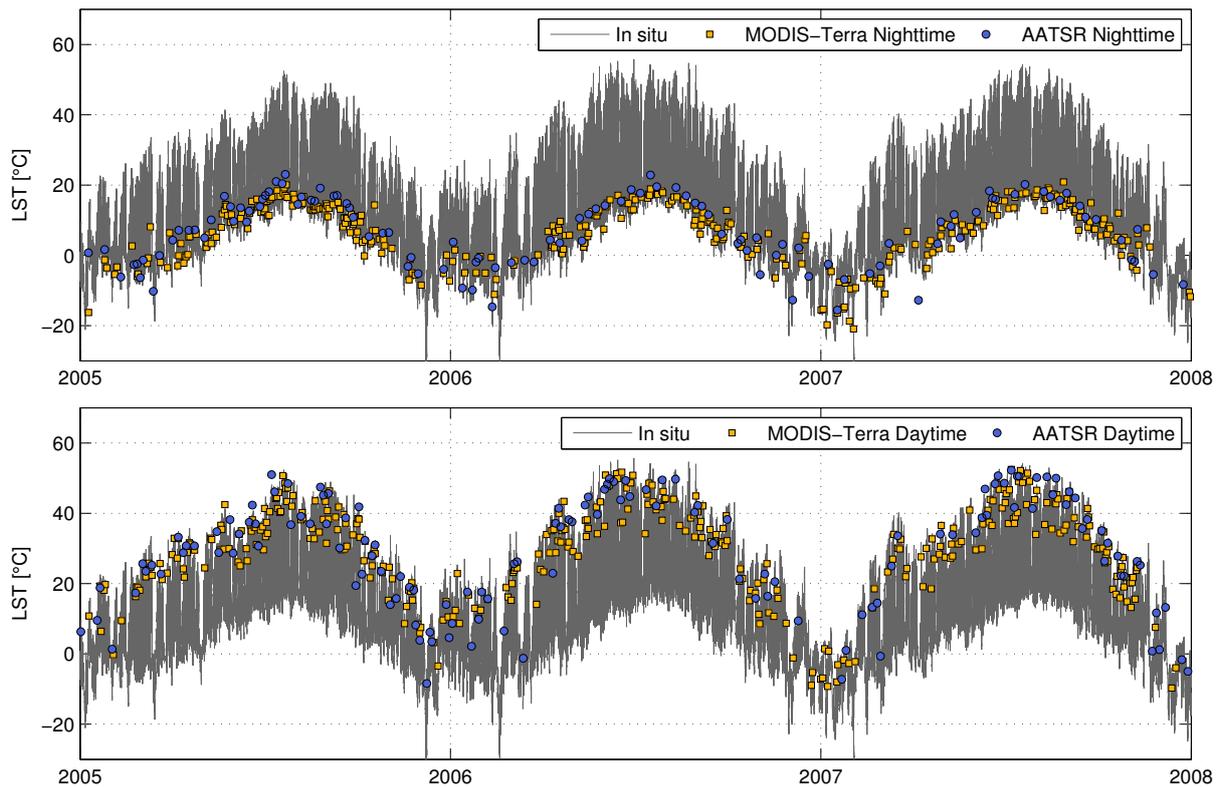


Figure 13 – Inter-comparison of the full nighttime (top) and daytime (bottom) LST time series derived from AATSR versus MODIS-Terra for the entire WACMOS-ET study period at the Table Mountain station. For reference the continuous in situ LST time series measured at the Table Mountain station is also given.

Figures 14 through 17 show the difference images for daytime and nighttime data for 15 July 2007 and 15 December 2007, respectively. These images were constructed by first mosaicking the individual sinusoidal tiles from both MODIS MOD11 LST and AATSR LST into a global grid. Subsequently the global AATSR mosaic was subtracted from the MODIS mosaic at all grid cells where the orbit from both satellite instruments overlapped. With the exception of Figure 17, which shows higher values for AATSR than for MODIS globally, all other Figures indicate both positive and negative differences. For daytime data on 15 July 2007 the positive difference (i.e. AATSR is higher than MODIS) interestingly occurs nearly exclusively in the southern hemisphere. For nighttime data of the same day, however, some positive discrepancies can be found mostly over Russia. The positive differences for daytime data on 15 December 2007 on the other hand are found across the globe with no specific spatial clustering.

Figures 18 through 21 show the corresponding scatter plots of MODIS LST versus AATSR LST for both daytime and nighttime data on 15 July 2007 and 15 December 2007. The scatter shown for daytime data on 15 July 2007 (Figure 18) shows as expected for daytime data a relatively large amount of scatter. It should be noted that the scatter is relatively low for temperatures of around 10 °C and less, but quite large for higher temperatures up to 60 °C. The overall bias was found to be 1.7 °C, indicating that AATSR LST retrievals are on average higher than those from MODIS. The RMSE as a measure of overall correspondence was found to be relatively high with 3.4 °C but the coefficient of determination indicated a quite good agreement of a linear trend model ($R^2 = 0.95$). The nighttime plot for the same day (Figure 19) shows a reduced amount of scatter with an RMSE of only 2.7 °C, but the bias between the two data sources is even increased with a value of 2.0 °C.

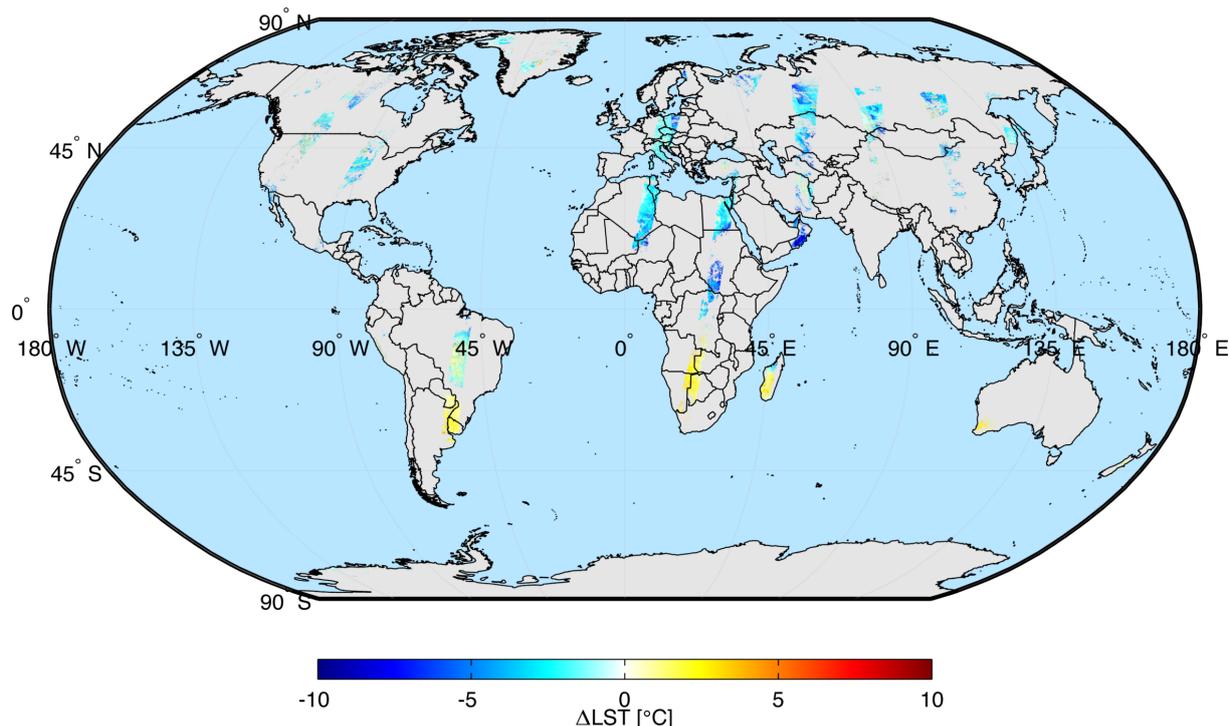


Figure 14 – Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for daytime data for 15 July 2007.

The scatterplot for daytime data on 15 December 2007 shows a significantly reduced amount of bias with a value of only 0.5 °C. The reason for this reduction is not entirely clear but it is likely related to the different areas on the globe at which both MODIS and AATSR data are available for a certain day. Since for 15 December 2007 the total number of matchup pixels was significantly reduced and the available overlap was primarily in North Africa, the statistics derived for this day are likely not as representative as those derived for 15 July 2007 for which much larger areas of overlap were available (see Figures 14 and 15). The scatterplot for nighttime data on 15 December 2007 (Figure 21) again shows a higher bias between the two data sources with a value of 2.0 °C. In fact, also the random error with a standard deviation of 1.9 °C and an RMSE of 2.8 °C was at a very similar magnitude as for the July comparison. This indicates that the statistics for nighttime data appear to be robust and do not seem to vary too much from day to day.

The bias for all four cases studied here indicates the LST retrievals from AATSR are higher than those from MODIS MOD11 by between 0.5 °C and 2 °C, with three out of four showing a value close to the latter. Keeping in mind the results from the previous section which showed that the MODIS product itself showed a significant low bias at least at the studied in situ sites, this can be considered as somewhat encouraging in the sense that AATSR does compensate for the (erroneous) bias from MOD11 and thus seems to provide a better estimate of LST. While similar biases for the MODIS MOD11 product have been reported before (Wang et al., 2008), more research will be necessary to explain this discrepancy. It should also be noted that a new LST and emissivity product (MOD21) is planned to be released as part of MODIS Collection 6 and it is likely that this product will mitigate the issues with negative biases (Hulley et al., 2014).

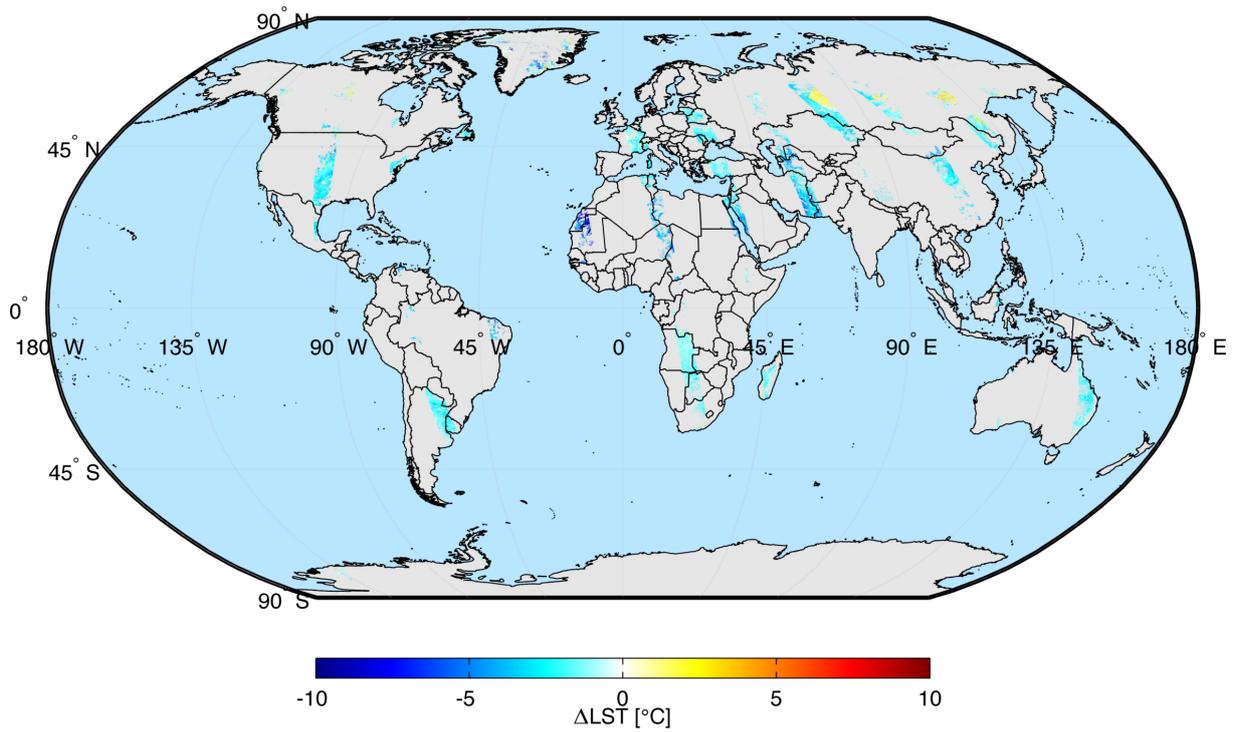


Figure 15 – Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for nighttime data for 15 July 2007.

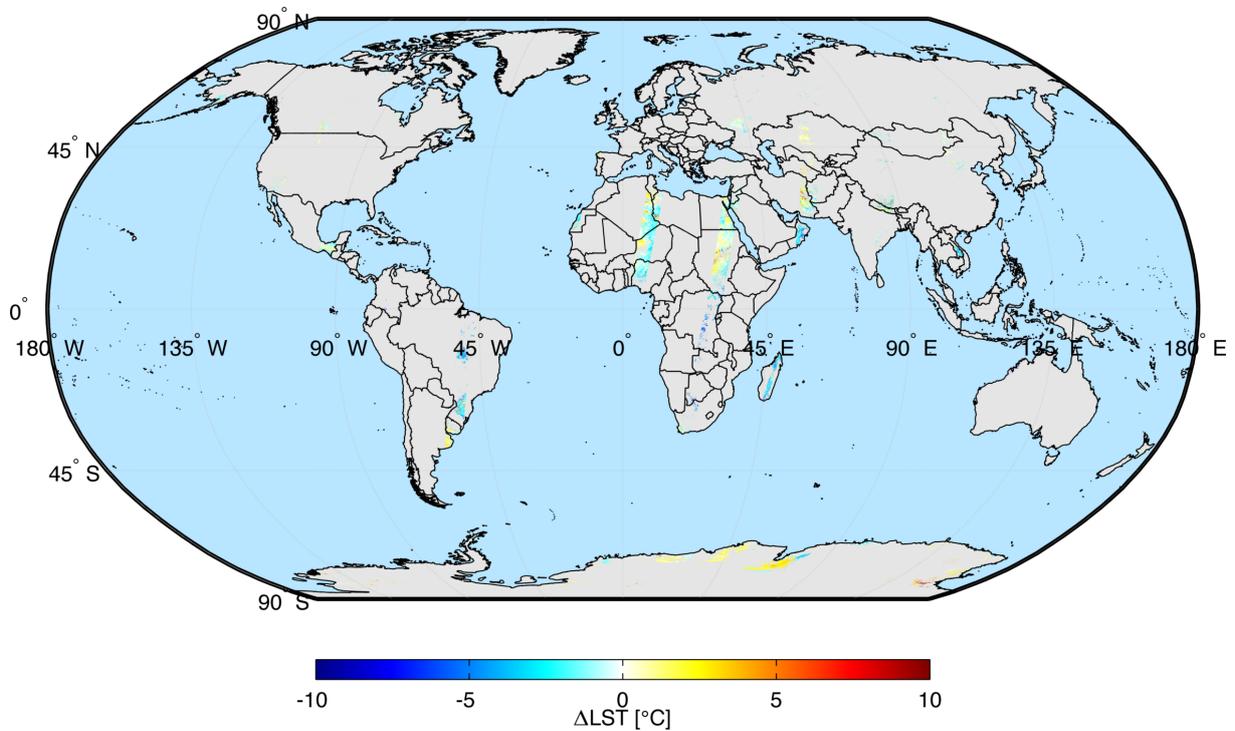


Figure 16 – Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for daytime data for 15 December 2007.

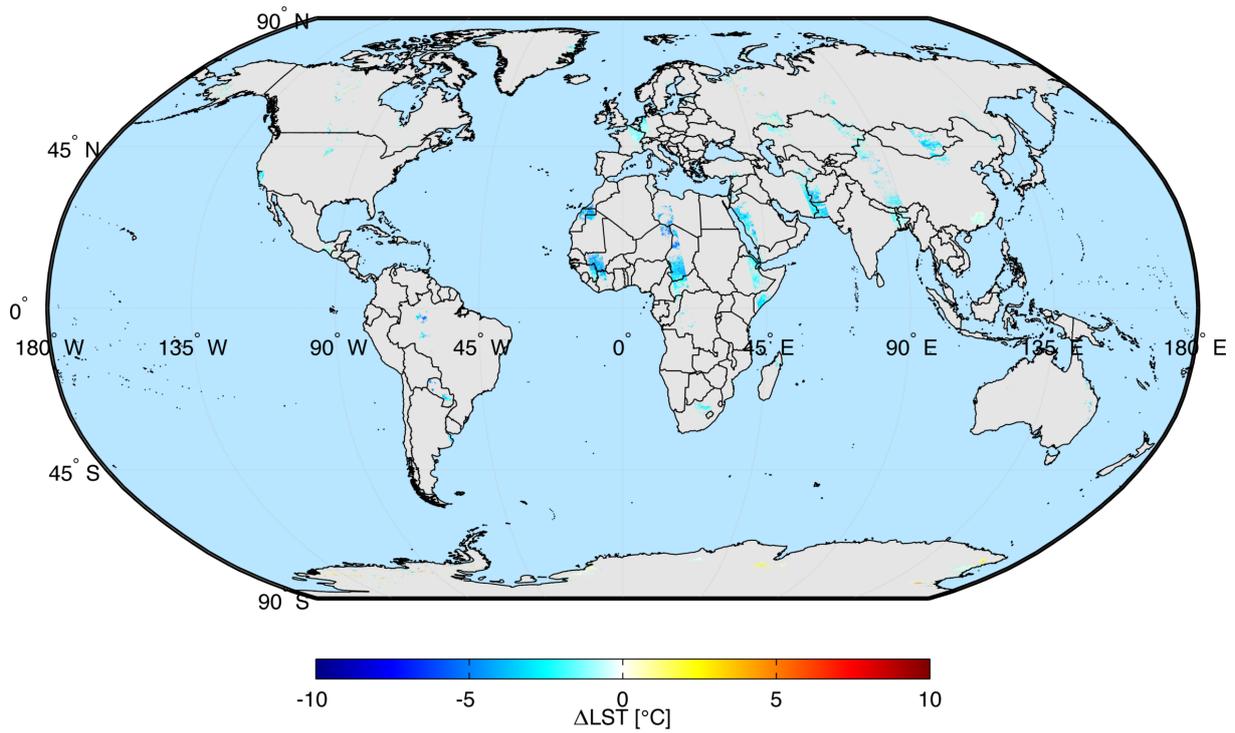


Figure 17 – Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and AATSR-derived LST for nighttime data for 15 December 2007.

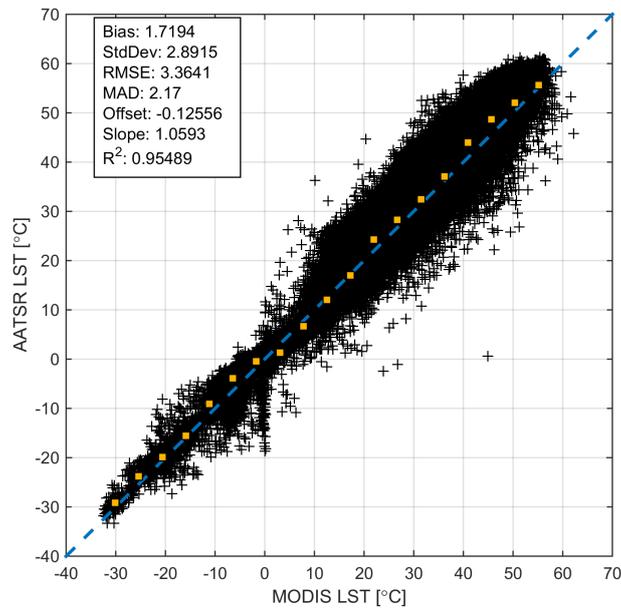


Figure 18 – Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for daytime data for 15 July 2007. Yellow markers indicate the median value for various temperature classes.

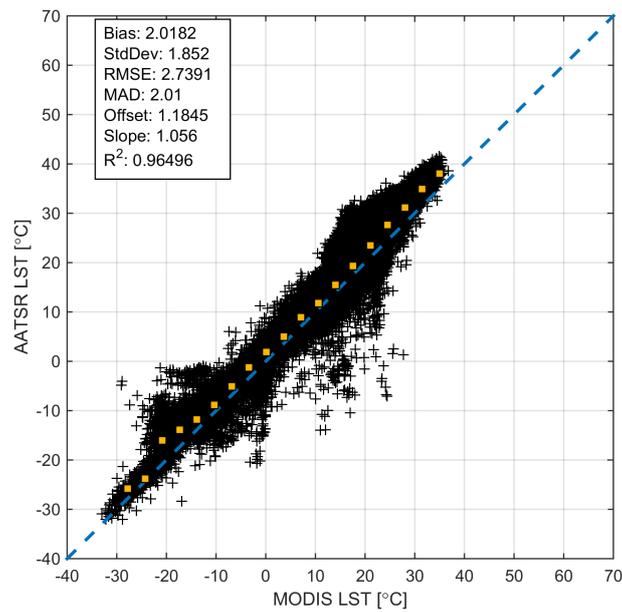


Figure 19 – Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for nighttime data for 15 July 2007. Yellow markers indicate the median value for various temperature classes.

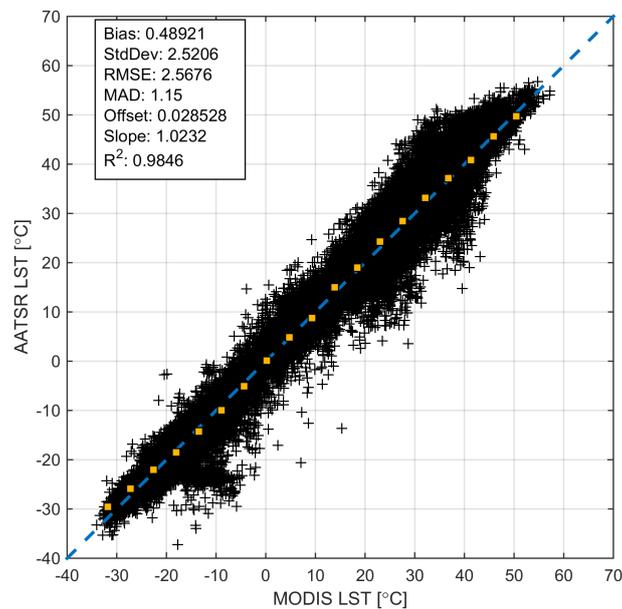


Figure 20 – Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for daytime data for 15 December 2007. Yellow markers indicate the median value for various temperature classes.

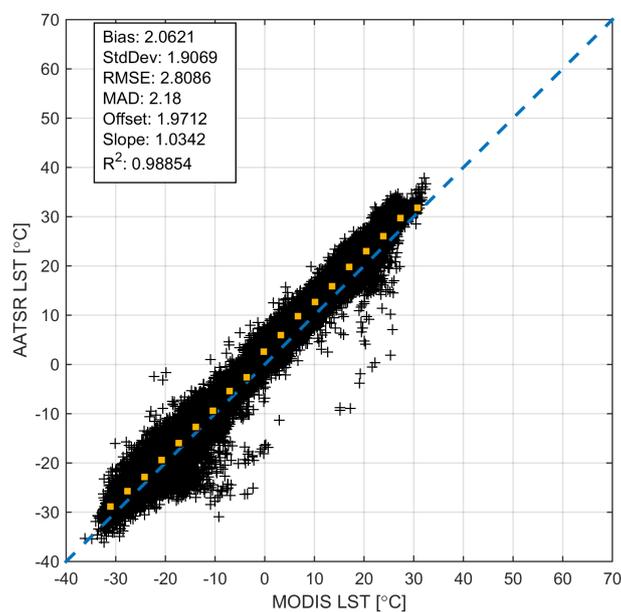


Figure 21 – Scatterplot showing the relationship between MODIS-based LST and AATSR-derived LST for nighttime data for 15 December 2007. Yellow markers indicate the median value for various temperature classes.

3.2.2 Comparison with GlobTemperature

The AATSR LST product derived within the framework of WACMOS-ET was further evaluated against the AATSR LST product generated as part of the ESA-funded GlobTemperature project (see www.globtemperature.info). The GlobTemperature product uses an algorithm based on the one developed by Prata (2002) but is improved in terms of the spatial resolution of the auxiliary datasets such as fractional vegetation cover and global biome distribution.

Figure 22 shows the time series of the two LST products for the year 2007 at 9 validation stations. Both daytime and nighttime retrievals are shown. The in situ-observed LST is also given as a reference. Qualitatively the LST retrievals from the two products agree quite well with each other and the in situ time series for the majority of stations. The temperature differences between the two products are quite small in most cases and their time series follow the in situ observations very closely.

Nonetheless, there are significant differences when a more in-depth comparison with the in situ data is carried out. Figure 23 shows the time series of the differences between the daytime/nighttime retrievals of both products and the corresponding in situ observations. One of the most obvious patterns is the impact of the previously mentioned nighttime cloudmasking issues with the WACMOS-ET AATSR product, which show up in this Figure as positive outliers. This can be seen for example at the Bondville, Desert Rock, Fort Peck, and Sioux Falls stations. Furthermore, a few negative outliers in the WACMOS-ET AATSR product exist for daytime data, particularly at the Bondville station.

In order to better quantify the discrepancies, Table 8 shows the validation results of the WACMOS-ET AATSR product at the nine in situ stations for the year 2007. Compare this to the corresponding results from the GlobTemperature product presented in Table 9. For the period and the locations investigated here the GlobTemperature LST product outperforms the WACMOS-ET LST product for daytime retrievals but the opposite is true for nighttime retrievals (as indicated by the RMSE). The average nighttime biases are 0.03 °C and 0.18 °C for WACMOS-ET and GlobTemperature, respectively. For daytime data, the biases are -0.70 °C and -0.50 °C, respectively. While the biases are relatively similar, particularly considering the relatively low number of retrievals from which these statistics were calculated, the standard deviations are somewhat higher for the WACMOS-ET product. Particularly for nighttime data, this discrepancy can be to some extent explained by the cloudmasking issues mentioned earlier, which resulted in several outliers (see Figure 23). Once this issue is taken care of, the two products are likely to have quite similar accuracies.

It should be noted that this comparison was carried out only over a 12-month timeframe and thus the statistics are based on a relatively low number of individual retrievals. Nonetheless they should be able to give a reasonable indication of the relative product accuracies.

As the differences between the WACMOS-ET AATSR product and the GlobTemperature AATSR product are to some extent due to differences in cloud masking, the comparison was repeated and the statistics for the GlobTemperature were re-calculated using the same cloud-mask as used by WACMOS-ET. Table 10 shows the results. In addition, Figures 24 and 25 show the time series of LST and differences between in situ and LST for both products, now using the same cloud masks.

Finally, overall summary statistics for both products were computed together over all in situ sites for both products, only classified by daytime and nighttime retrievals. The results are shown in Table 11. Overall, the WACMOS-ET AATSR product provides slightly more accurate results than the GlobTemperature AATSR LST product when the same cloudmask is used. The RMSE as a measure of overall accuracy including both systematic and random errors is lower for the

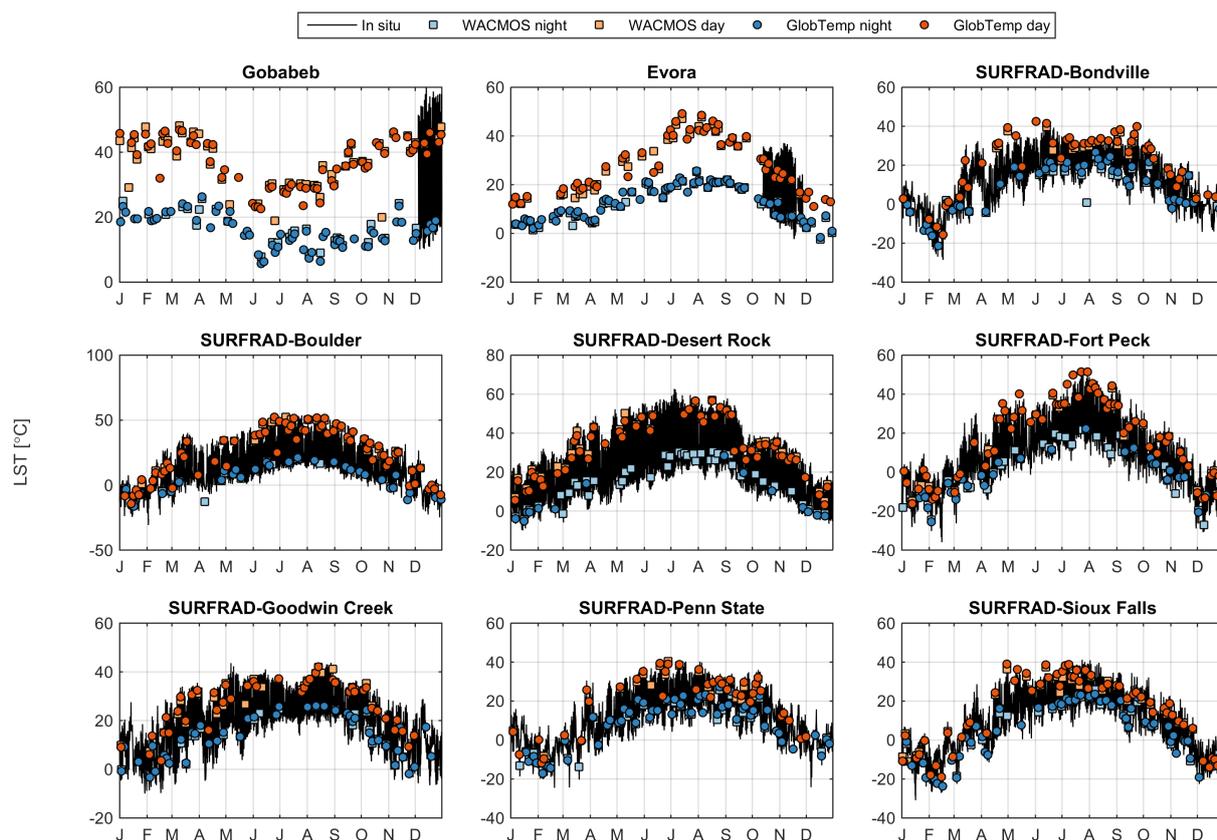


Figure 22 – Comparison of the GlobTemperature AATSR LST product versus the WACMOS-ET derived AATSR LST product, shown here for the year 2007 and nine locations with in situ observations.

Table 8 – Summary statistics of the comparison of the WACMOS-ET derived AATSR LST product, calculated for the year 2007 at nine locations with in situ observations.

St #	Station	N	Nighttime			N	Daytime		
			Bias	StdDev	RMSE		Bias	StdDev	RMSE
1	Gobabeb	0				5	-1.50	0.76	1.64
2	Evora	7	0.71	0.75	1.00	5	4.04	1.88	4.38
3	Bondville	36	-0.02	4.31	4.25	22	-2.21	3.22	3.84
4	Boulder	27	0.97	2.15	2.32	36	-3.11	2.81	4.16
5	Desert Rock	47	0.46	0.96	1.06	41	-1.80	4.09	4.42
6	Fort Peck	44	0.23	1.88	1.88	29	-1.34	2.87	3.13
7	Goodwin Creek	29	-0.99	1.68	1.93	35	-0.62	2.04	2.10
8	Penn State	45	-0.95	1.76	1.98	25	-0.18	2.65	2.60
9	Sioux Falls	36	-0.19	1.52	1.51	35	0.42	2.95	2.94
Average		30	0.03	1.88	1.99	26	-0.70	2.58	3.25

WACMOS-ET product both for daytime and nighttime retrievals. It should be noted, however, that in the case of nighttime data the difference in RMSE is very small, but the bias is reduced significantly from 0.15 °C to 0.06 °C. For daytime data the reduction in bias is very small, but the WACMOS-ET AATSR product exhibits an RMSE which is lower than that of the GlobTemperature product by more than 0.5 °C.

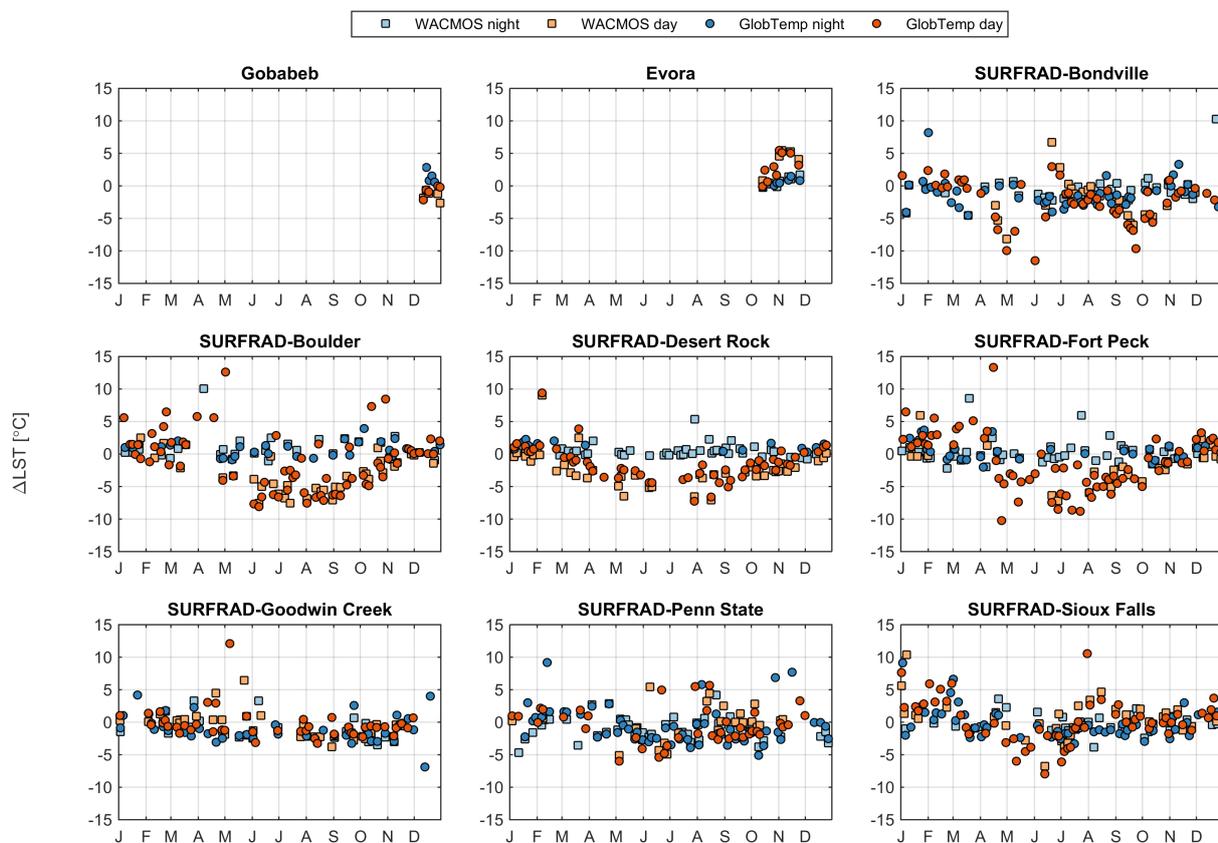


Figure 23 – Differences between the GlobTemperature AATSR LST product and the WACMOS-ET derived AATSR LST product against the corresponding in situ LST observations, shown here for the year 2007 at nine locations.

Table 9 – Summary statistics of the comparison of the GlobTemperature AATSR LST product, calculated for the year 2007 at nine locations with in situ observations.

St #	Station	N	Nighttime			Daytime			
			Bias	StdDev	RMSE	N	Bias	StdDev	RMSE
1	Gobabeb	4	1.43	1.01	1.67	5	-0.79	0.85	1.10
2	Evora	8	0.53	0.53	0.73	9	2.93	1.99	3.48
3	Bondville	47	-1.17	2.15	2.42	49	-2.43	3.28	4.05
4	Boulder	30	0.80	1.12	1.36	67	-1.13	4.35	4.47
5	Desert Rock	14	1.25	0.60	1.38	54	-1.10	3.70	3.83
6	Fort Peck	29	0.41	1.50	1.53	68	-1.37	4.30	4.48
7	Goodwin Creek	43	-1.01	2.03	2.24	44	-0.32	2.35	2.34
8	Penn State	57	-0.57	2.87	2.90	39	-0.51	2.74	2.75
9	Sioux Falls	55	-0.07	2.36	2.34	57	0.18	3.38	3.35
Average		32	0.18	1.57	1.84	44	-0.50	2.99	3.32

Table 10 – Same as Table 9 but using the same cloud-mask as used for the WACMOS-ET AATSR product.

St #	Station	N	Nighttime			Daytime			
			Bias	StdDev	RMSE	N	Bias	StdDev	RMSE
1	Gobabeb	1	1.52	0.00	1.52	5	-0.79	0.85	1.10
2	Evora	8	0.53	0.53	0.73	3	3.36	2.97	4.14
3	Bondville	43	-0.79	4.13	4.15	34	-2.88	4.71	5.46
4	Boulder	32	1.00	1.80	2.04	49	-2.62	3.47	4.32
5	Desert Rock	57	1.60	0.80	1.78	51	-0.77	4.92	4.93
6	Fort Peck	49	0.08	1.32	1.31	33	-0.91	3.68	3.73
7	Goodwin Creek	35	-1.35	1.46	1.97	40	-0.39	2.21	2.21
8	Penn State	52	-1.38	1.77	2.23	31	-1.01	2.62	2.77
9	Sioux Falls	46	-0.63	1.56	1.66	44	-0.28	3.38	3.35
Average		36	0.06	1.48	1.93	32	-0.70	3.20	3.56

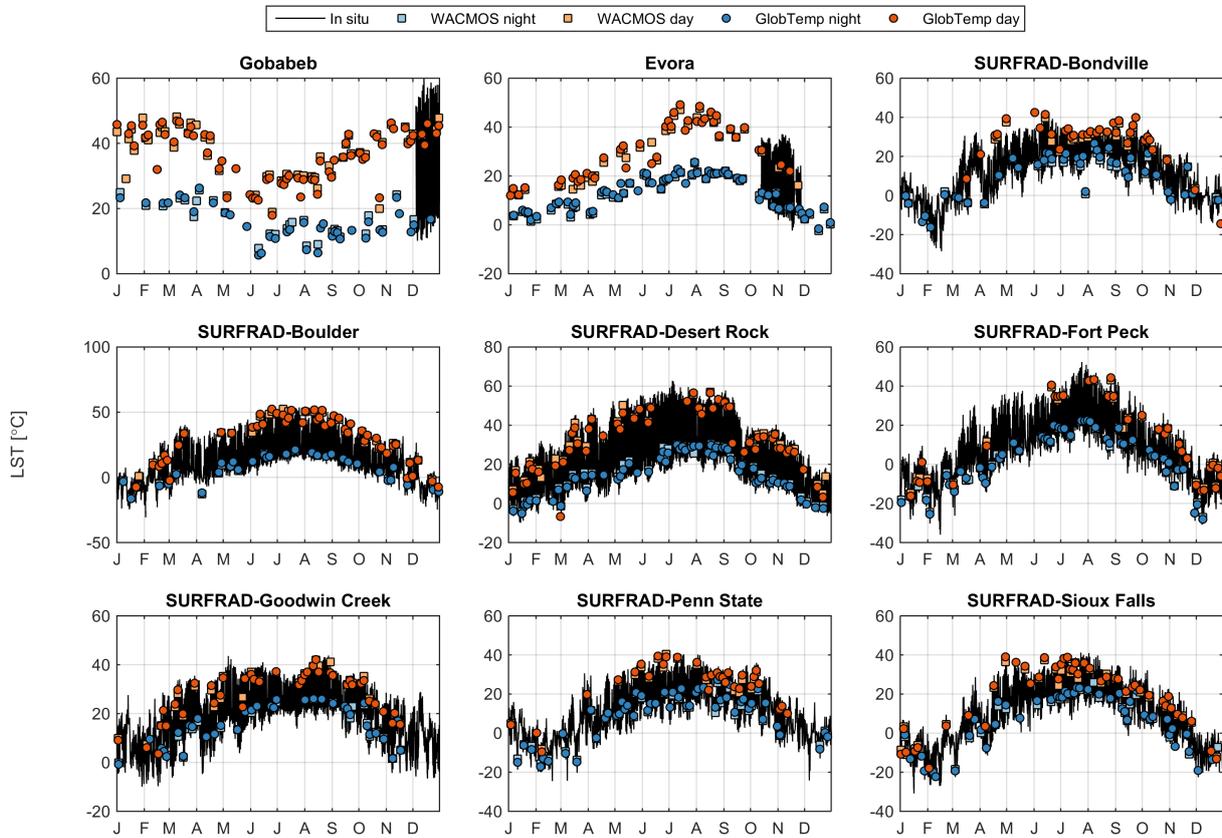


Figure 24 – As Figure 22 but using the original L1 cloud mask for the GlobTemperature product.

Table 11 – Overall summary statistics of all nighttime and daytime matchups for 2007 at all in situ sites. Note that the same cloudmask was used for both products.

Product	Bias	StdDev	RMSE	MAD	Offset	Slope	R ²
WACMOS Night	0.06	2.26	2.25	0.81	0.34	0.97	0.97
GlobTemp Night	0.15	2.27	2.27	1.38	0.19	1.00	0.97
WACMOS Day	1.17	3.25	3.45	2.17	-0.76	1.08	0.96
GlobTemp Day	1.20	3.81	3.99	1.95	-1.52	1.11	0.96

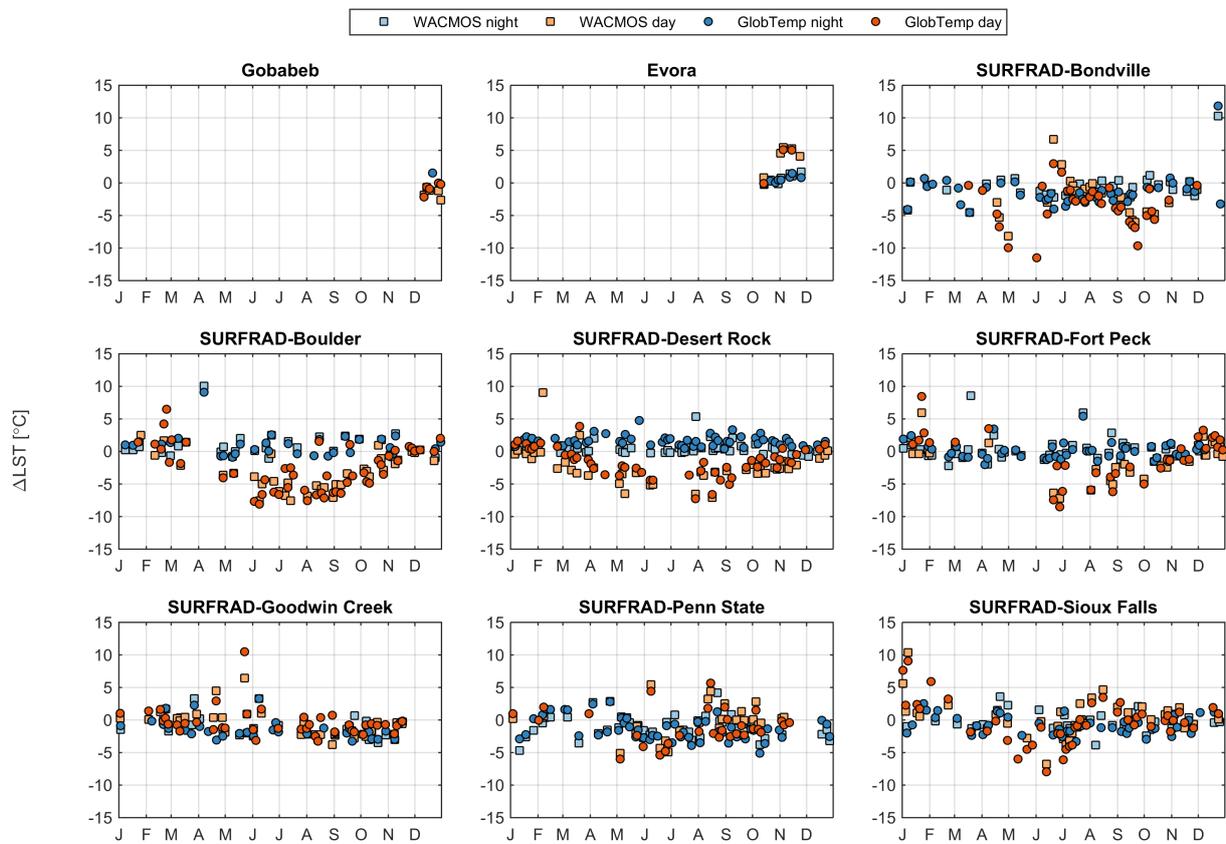


Figure 25 – As Figure 23 but using the original L1 cloud mask for the GlobTemperature product.

4 Validation of MTSAT

4.1 Validation with in situ observations

Only one of the available in situ LST stations was available in the coverage area of MTSAT. This station from the ARM network is located near the town Darwin in Northern Australia. This shortage of stations has already been noted in the past by Renzullo (2009).

It should be noted that the Darwin ARM station is far from ideal for validation of LST. The landscape around the site is highly heterogeneous and thus emissivity around the station itself is quite variable. Furthermore, the station is located relatively close to the coastline, which means that due to the relatively coarse spatial resolution of MTSAT ($0.04^\circ \times 0.04^\circ$), the MTSAT pixel located right above the station was partially covered by ocean. This renders the comparison invalid as the MTSAT LST will be strongly affected by the emissivity of the ocean, which is quite different than that of land surface. In order to obtain a more meaningful comparison, we chose a neighbouring MTSAT pixel located to the southeast of the original location and extracted a time series there. The chosen site is quite homogeneous and thus more suitable for comparing the MTSAT LST time series. However, this location is a bit further inland than the Darwin in situ site and thus the characteristics of the temperature time series are somewhat different. They have a more continental character and show lower nighttime and higher daytime LSTs than the Darwin station itself. Nonetheless approximate qualitative comparisons can still be carried out.

Despite these shortcomings, the Darwin ARM station was at the time of the WACMOS-ET study period the only station delivering publicly available and usable in situ LST observations within the MTSAT coverage area. For this reason, the data was used here for MTSAT validation.

Figure 26 shows a comparison of the time series of both the in situ observation at the Darwin ARM station and the MTSAT LST time series. The LST for the in situ data at this tropical station ranges from approximately 10 °C in the nighttime in the winter to nearly 50 °C in the daytime in the summer months. The MTSAT LST follows the seasonal cycle between wet and dry seasons quite well, however as the time series had to be extracted at a slightly more inland location as described above, this site shows a slightly more continental behavior and therefore exhibits slightly lower minimum LSTs and in some cases slightly higher maximum LST values.

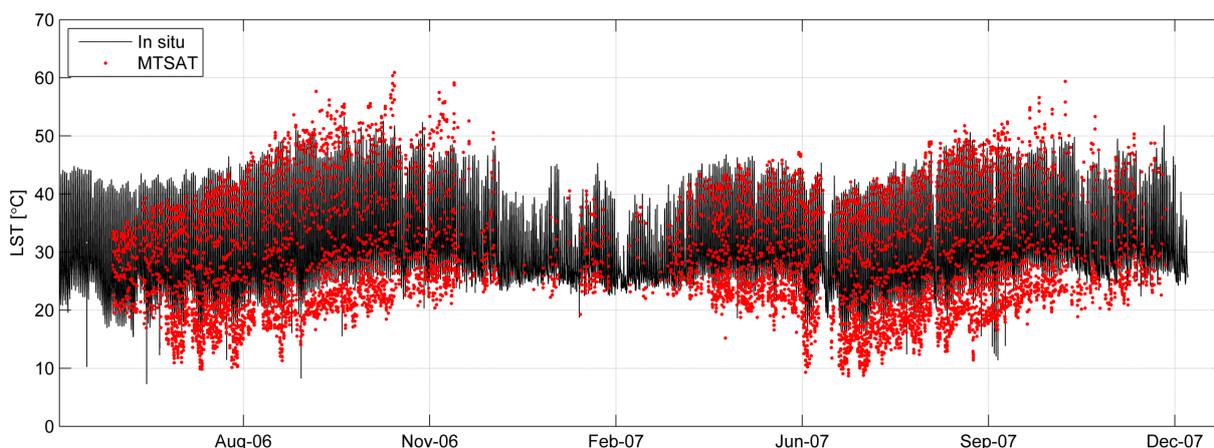


Figure 26 – Time series of in situ LST observations at the Darwin station and MTSAT retrievals at a nearby grid cell. Note that the grid cell directly located over the in situ station could not be used as it was partially covering ocean.

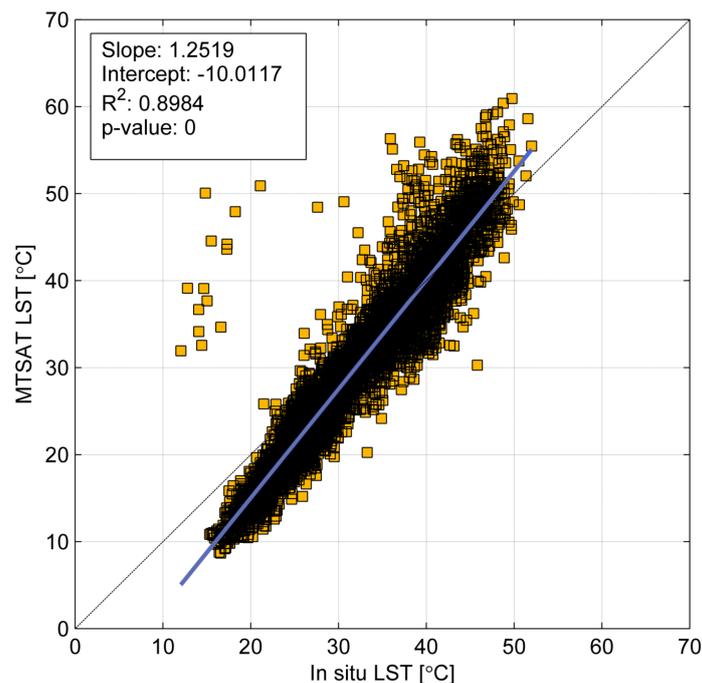


Figure 27 – Scatterplot of LST ground observations at Darwin, Australia, against LST retrieved from MTSAT over a nearby location. Note that the grid cell directly located over the in situ station could not be used as it was partially covering ocean.

Figure 27 shows the two time series in a scatter plot, which is better able to indicate their relationship. It shows a quite strong relationship with an R^2 value of nearly 0.9. However, it is also obvious that the linear regression does not follow exactly the 1 to 1 line but has a slope which is slightly steeper than 1, namely 1.25. As such it shows that MTSAT underestimates low in situ LST values and overestimates high in situ LST values. However it should be noted that this clearly is indicative of the distance between the in situ site and the slightly more inland location of the extracted MTSAT time series and would likely not occur if the in situ site were located in a more suitable, i.e. homogeneous location, which would allow for extracting MTSAT data right over the station. A dedicated in situ validation site for LST, as it is available for example for Africa with Gobabeb, would be extremely useful in the coverage area of MTSAT. Such a dedicated station has been operated in Australia in the past (Prata, 1994a,b, 2003) but the data collection there stopped before the WACMOS-ET study period and thus this otherwise quite useful dataset could not be used here for MTSAT validation.

Nonetheless, despite its shortcomings due to lack of suitable in situ data, this comparison shows that MTSAT has great potential for providing good LST products, but a true absolute validation of MTSAT LST against in situ data will only be possible with a more suitably located in situ station.

A somewhat optimistic outlook can also be gained by Figure 28, which shows time series of two days of data in July 2007. In this case the MTSAT LSTs follow the observed in situ LST remarkably well. As such, while the accuracy of the derived MTSAT LST might not reach the same accuracy of LST derived from other instruments such as AATSR or MODIS, it is still very useful for obtaining information about the diurnal cycle of LST in the region of Asia and Oceania.

4.2 MTSAT Validation against independent EO data

In addition to a validation against ground-based in situ data, a comprehensive inter-comparison of the MTSAT LST dataset produced within the framework of WACMSOS-ET against an inde-

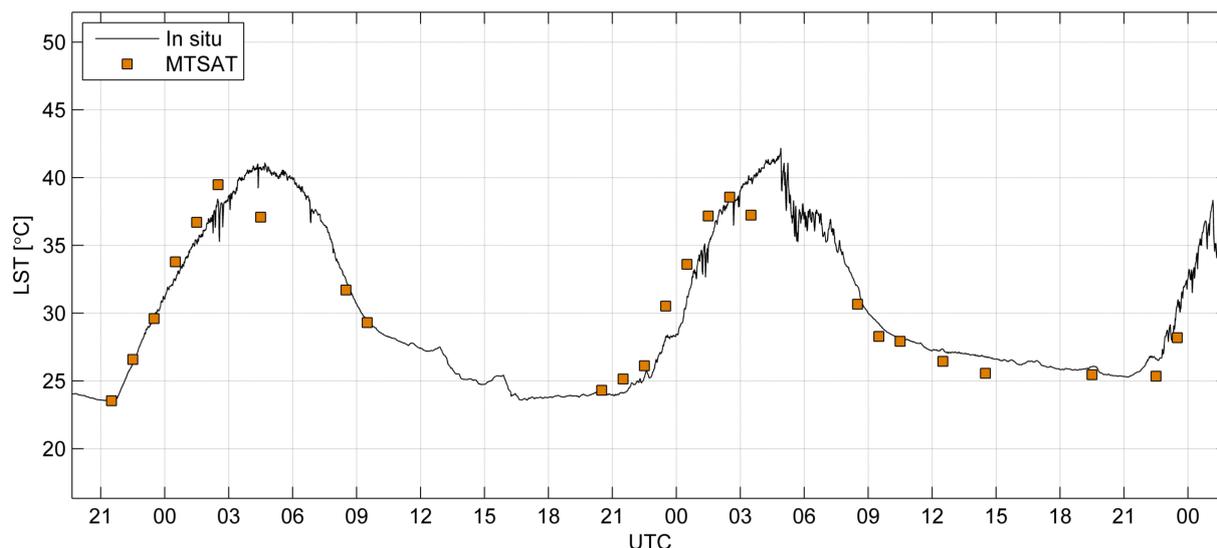


Figure 28 – Similar to Figure 26 but only showing two days in July 2007. On the scale of individual days MTSAT LST is able to adequately trace the diurnal temperature cycle even though systematic biases can be detected overall.

pendent remote sensing product was carried out. This comparison was particularly important as no suitable in situ was available to validate MTSAT LST. The reference product used here was the extensively validated MODIS Terra MOD11A1 product (Wan et al., 2002, 2004; Wan and Li, 2008; Wan, 2008; Coll et al., 2005, 2009; Galve et al., 2007). A similar comparison has been carried out in the past by Renzullo (2009).

This section is organized in three subsections. In the first one, the full global mosaic from MODIS is compared against a full MTSAT mosaic consisting of one single MTSAT image taken at the same time. Note that in this comparison there can be spatial variability in terms of observation time difference as the MTSAT image is consistent at a single moment in time whereas the MODIS global mosaic is composited from multiple satellite overpasses. For this reason the second subsection performs a tile-by-tile comparison between MTSAT and MODIS which allowed to use the best matchup in time for each individual tile. Finally, the third subsection investigates the difference between MTSAT LST and MODIS LST characterized in terms of land cover, while the fourth subsection briefly investigates the impact of the zenith view angle.

4.2.1 Inter-comparison of full mosaic

The inter-sensor validation was first carried out by comparing full mosaics for the entire globe and the Australia region derived from MODIS and MTSAT, respectively. For this purpose, two seasonally different days in 2007 were selected, namely 15 July 2007 and 15 December 2007. For the synoptic mosaic of MTSAT a time was selected that appeared to be the best compromise with respect to the individually best-matching times at the tile level. For daytime data 04:30 UTC was selected to compute the MTSAT mosaic, whereas for nighttime data 15:30 UTC was chosen after careful investigation of MODIS/MTSAT matchups times for the individual tiles. Note that this will introduce slightly larger biases in some tiles which are close to an earlier or later MODIS overpass, but this was considered acceptable in order to obtain spatial consistency within the MTSAT mosaic. Note that Section 4.2.2 performs a tile-by-tile comparison to overcome the issue. The analysis was carried out both for daytime and nighttime data. It should also be noted that the daytime differences in general, but in particular in this comparison, will always be higher than the nighttime differences due to the large daily diurnal cycle of temperature.

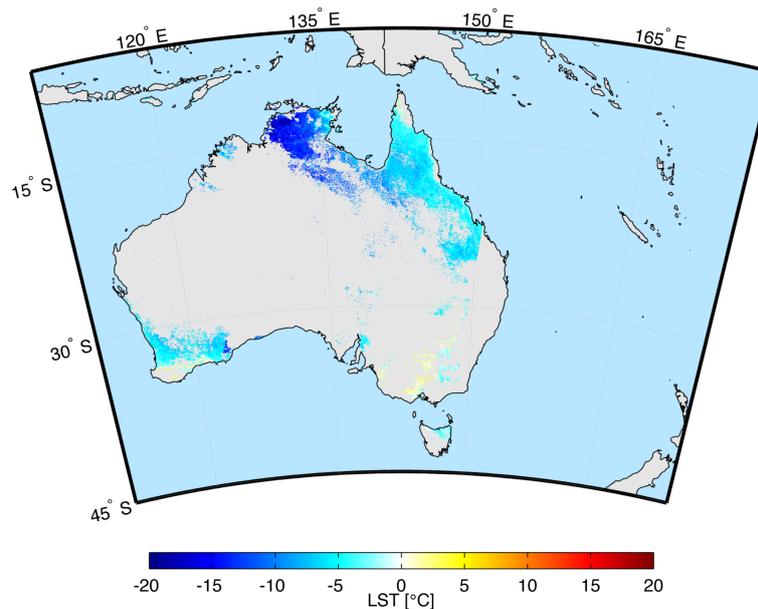


Figure 29 – Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for daytime data on 15 July 2007.

Similar differences in observation time will therefore usually result in higher inter-sensor biases for daytime comparisons than for nighttime comparisons.

Figure 29 shows the discrepancy between MOD11 and MTSAT for the Australia region on 15 July 2007 for daytime data. Nearly all across Australia, with an exception in the states of Victoria and New South Wales, the MTSAT LSTs are higher than the MODIS MOD11 LSTs by several degrees. In the northern half of the Northern Territory the discrepancies even reach very high values of more than 10 °C. Based on the results of the previously shown comparison of the MOD11 product with in situ data it appears as if the MOD11 product tends to significantly underestimate daytime LSTs by around 1-2 °C. This could explain some of the higher MTSAT LSTs to a certain extent, but the biases found for daytime data particularly in the Darwin region of North Australia are clearly too high for this being the sole explanation. The strong discrepancies for daytime data do not appear to be very closely linked to land cover. Most of the northern areas of Australia are covered by trees ranging from sparse to closed.

Figure 30 shows the corresponding image for nighttime data. Here it is obvious that nearly throughout the entire Australian continent MOD11 exhibits higher LST values than MTSAT. While the discrepancies are not very large and only on the order of 1-2 °C, they are fairly spatially consistent with only a few very small areas in the north showing the opposite signal.

Figures 31 and 32 show the same respective daytime and nighttime images for the summer season on December 2007. In contrast to the wintertime images, the number of matching pixels between MTSAT and MODIS is significantly reduced here such that no firm conclusions can be drawn based on this data. However, the general tendency shows the opposite behavior from the wintertime images - MTSAT provides higher LST values than MODIS for daytime data and lower temperatures than MODIS for nighttime data.

Figure 33 shows a scatterplot of the difference image between daytime MODIS and MTSAT LST for 15 July 2007, i.e. for southern hemisphere winter. As expected for daytime data, the plot indicates a strong positive bias for MTSAT of 6.7 °C with respect to MODIS. The bias is particularly high for warm temperatures of 25 °C and more. For lower temperatures between 0 and 20 °C the bias is relatively low. Both standard deviation and RMSE are quite high with values of 4.2 °C and 7.9 °C, respectively.

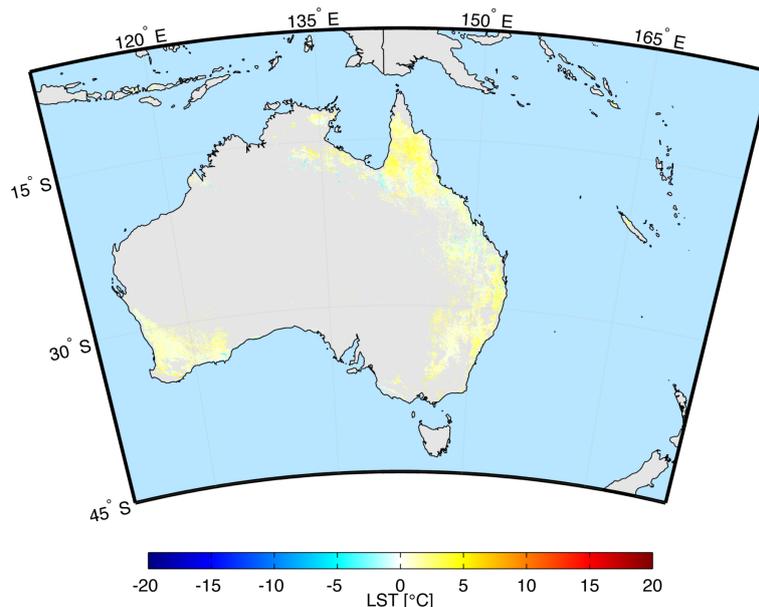


Figure 30 – Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for nighttime data on 15 July 2007.

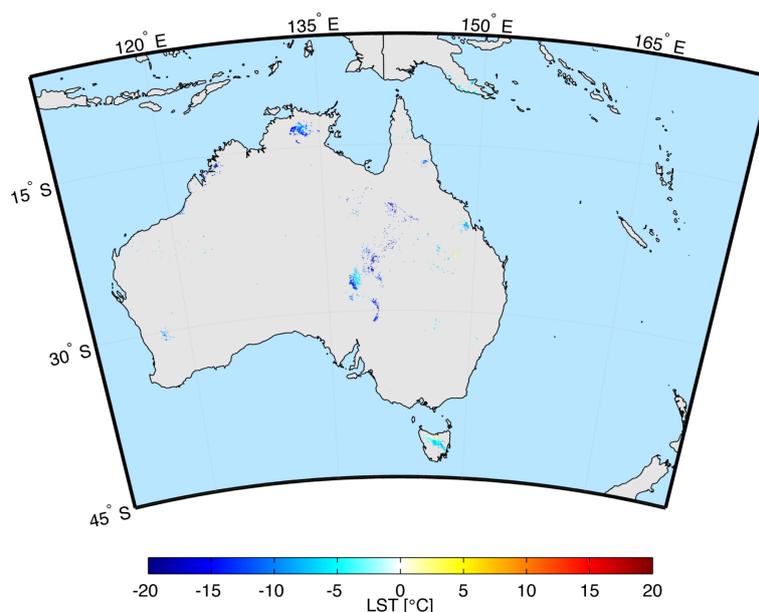


Figure 31 – Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for daytime data on 15 December 2007.

A much better relationship can be observed for nighttime data (Figure 34). Here, the scatter follows the 1:1 line much more accurately. The bias is quite low: MTSAT LST are on average -1 °C lower with respect to MODIS LST. Further, both standard deviation and RMSE are in very acceptable ranges with 1.4 °C and 1.8 °C, respectively. The coefficient of determination for the nighttime data on 15 July 2007 further is also very high with a value of $R^2 = 0.94$.

In the southern hemisphere summer, the situation is quite similar. Figure 35 shows the scatterplot for MODIS vs MTSAT LST daytime data for 15 December 2007. While the number of data points is slightly lower than for the southern hemisphere winter, the bias for daytime data is similarly high with 8.6 °C. Standard deviation and RMSE are also very high with values of 6.0 °C and 10.5 °C, respectively.

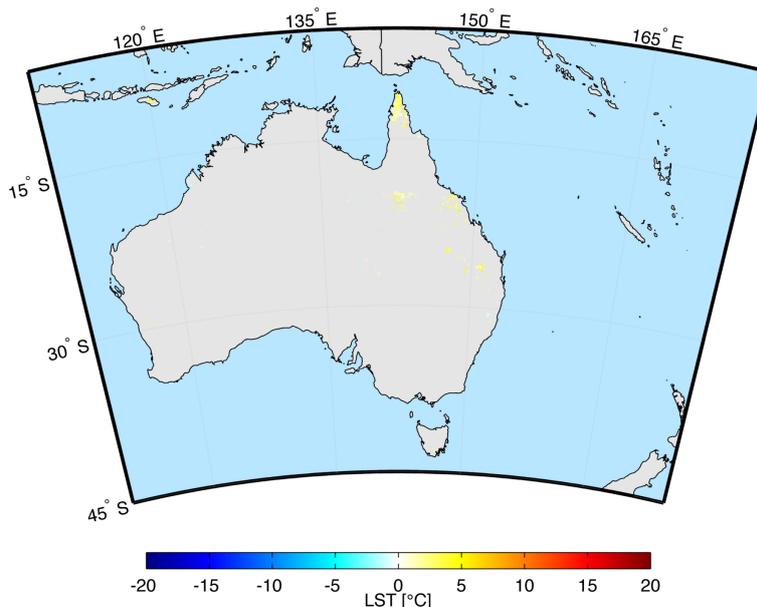


Figure 32 – Difference image of MODIS MOD11 LST minus MTSAT LST in the Australia region for nighttime data on 15 December 2007.

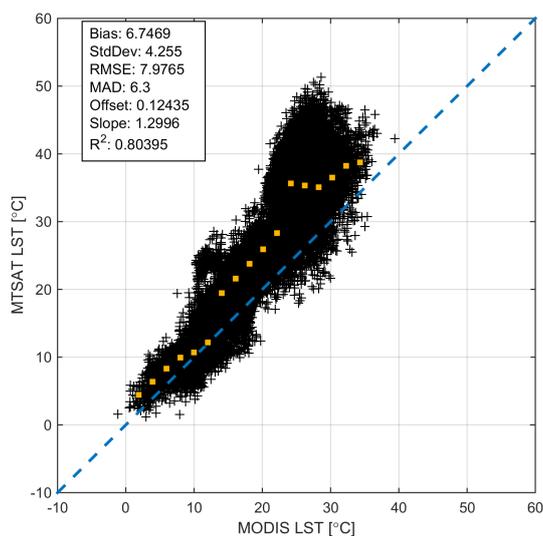


Figure 33 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 July 2007. Note that the matchups were derived from a full global mosaic of both MODIS and MTSAT data. For the latter the observations at 04:30 UTC were used. Yellow markers indicate the median value for various temperature classes.

Finally, Figure 36 shows the scatterplot for nighttime data on 15 December 2007. Again, the relationship is much improved over the daytime data but it does not quite reach the accuracy of nighttime comparison in southern hemisphere winter. The bias has a slightly higher absolute value with -1.5 °C , and the standard deviation and RMSE are slightly increased with values of 2.1 °C and 2.6 °C , respectively. Nonetheless these values are still quite reasonable and mostly agree with similar results reported by Renzullo (2009).

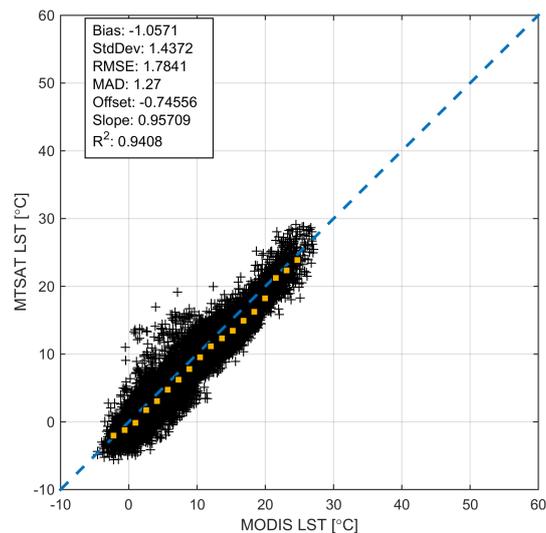


Figure 34 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 July 2007. Note that the matchups were derived from a full global mosaic of both MODIS and MTSAT data. For the latter the observations at 15:30 UTC were used. Yellow markers indicate the median value for various temperature classes.

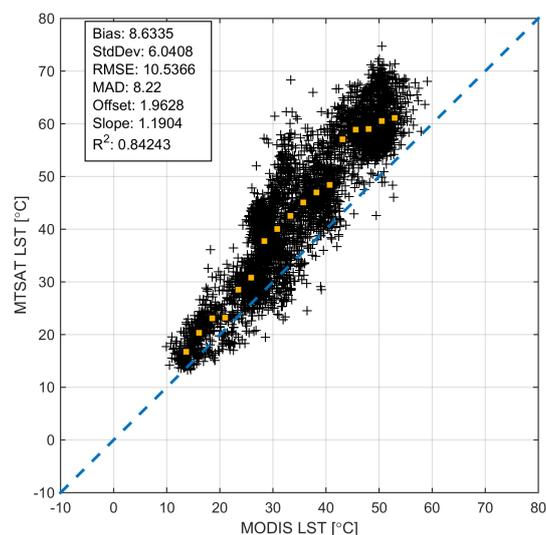


Figure 35 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 December 2007. Note that the matchups were derived from a full global mosaic of both MODIS and MTSAT data. For the latter the observations at 04:30 UTC. Yellow markers indicate the median value for various temperature classes.

4.2.2 Tile-by-tile inter-comparison of full global mosaic

In addition to the full mosaic comparison reported on in Section 4.2.1, a tile-by-tile comparison was carried out. This method has the advantage that the best MODIS overpass can be computed individually for each sinusoidal tile. As this reduces the number of pixels with large differences in observations between MTSAT and MODIS, this should reduce the overall error and make the comparison more robust. Again, the inter-comparison is carried out for one day in southern-hemisphere winter (15 July 2007) and one in southern-hemisphere summer (15 December 2007).

Table 12 shows the results for the daytime comparison between MTSAT LST and MODIS LST for 15 July 2007. Note that the biases in all the following tables were calculated as MODIS minus

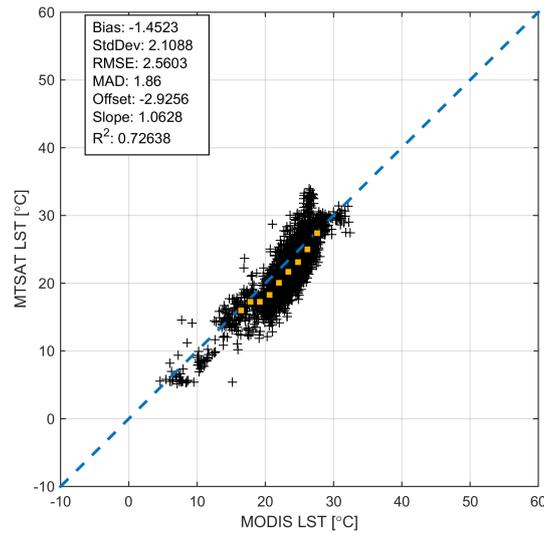


Figure 36 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 December 2007. Note that the matchups were derived from a full global mosaic of both MODIS and MTSAT data. For the latter the observations at 15:30 UTC were used. Yellow markers indicate the median value for various temperature classes.

Table 12 – Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for daytime data on 15 July 2007. Note that *Average* is just the mean of the results from each individual tile, i.e. there is no weighting by N, whereas *Overall* shows the statistics computed over the entire set of matching pixels independent of the respective tiles.

Tile No	Tile ID	N	Bias	StdDev	RMSE	MAD	Offset	Slope	R ²
1	h27v11	24	-2.4	0.7	2.5	2.4	15.8	0.38	0.27
2	h27v12	3925	-2.1	1.0	2.3	2.1	2.1	1.00	0.77
3	h28v09	0							
4	h28v10	0							
5	h28v11	503	-2.0	1.7	2.6	2.2	9.1	0.68	0.35
6	h28v12	4610	-3.6	1.9	4.1	3.2	14.1	0.40	0.22
7	h28v13	598	-2.6	1.4	3.0	2.6	2.5	1.02	0.64
8	h29v09	422	-1.6	2.9	3.3	2.5	6.6	0.83	0.33
9	h29v10	346	-4.5	2.1	4.9	4.7	8.2	0.88	0.58
10	h29v11	71	-4.0	2.0	4.5	4.1	13.2	0.49	0.19
11	h29v12	6122	-2.7	2.0	3.4	3.2	2.3	1.04	0.59
12	h29v13	3	5.1	0.2	5.1	5.0	-5.4	1.09	0.91
13	h30v09	93	-5.2	3.2	6.1	5.5	13.6	0.67	0.12
14	h30v10	9911	-8.0	2.6	8.4	8.0	15.3	0.70	0.25
15	h30v11	857	-5.0	2.0	5.3	5.2	-0.3	1.26	0.81
16	h30v12	2118	-3.7	2.0	4.2	4.1	3.5	1.01	0.67
17	h31v09	18	-0.1	1.1	1.1	1.0	1.8	0.94	0.19
18	h31v10	20576	-1.0	1.3	1.7	1.1	3.9	0.89	0.80
19	h31v11	9695	-5.9	1.7	6.1	5.9	8.4	0.88	0.75
20	h31v12	0							
21	h32v09	309	-4.3	2.7	5.1	4.9	5.8	0.94	0.36
22	h32v10	157	-1.1	2.0	2.3	1.6	8.1	0.71	0.37
23	h33v09	0							
24	h33v10	0							
25	h33v11	0							
Average		60358	-2.9	1.8	4.0	3.6	6.8	0.8	0.5
Overall		60358	-3.6	3.2	4.8	3.0	4.8	0.94	0.78

Table 13 – Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for nighttime data on 15 July 2007. Note that *Average* is just the mean of the results from each individual tile, i.e. there is no weighting by N, whereas *Overall* shows the statistics computed over the entire set of matching pixels independent of the respective tiles.

Tile No	Tile ID	N	Bias	StdDev	RMSE	MAD	Offset	Slope	R ²
1	h27v11	16	-0.6	1.6	1.7	1.3	6.2	0.25	0.07
2	h27v12	3928	0.4	1.3	1.4	0.8	1.0	0.74	0.45
3	h28v09	0							
4	h28v10	0							
5	h28v11	1133	0.0	1.4	1.4	0.8	0.9	0.87	0.66
6	h28v12	6222	0.3	1.9	1.9	0.8	0.9	0.71	0.26
7	h28v13	0							
8	h29v09	470	0.4	2.3	2.3	1.5	7.8	0.61	0.34
9	h29v10	436	-0.4	1.3	1.3	0.8	9.4	0.39	0.11
10	h29v11	59	0.1	1.7	1.7	1.1	1.2	0.75	0.74
11	h29v12	1936	0.5	1.4	1.5	0.9	-0.3	0.86	0.55
12	h29v13	1	0.7	0.0	0.7	0.7	0.0	0.75	
13	h30v09	9	-0.6	2.0	2.0	1.7	39.1	-0.90	0.28
14	h30v10	2374	-0.8	1.4	1.6	1.0	2.7	0.86	0.76
15	h30v11	488	-0.4	1.5	1.6	1.0	0.3	1.03	0.81
16	h30v12	7919	0.3	1.1	1.2	0.7	0.0	0.72	0.55
17	h31v09	11	0.0	1.0	0.9	0.6	12.1	0.48	0.17
18	h31v10	16751	0.3	1.2	1.2	0.8	1.4	0.88	0.85
19	h31v11	16432	-0.4	1.5	1.5	0.9	0.8	0.90	0.79
20	h31v12	367	0.9	1.3	1.6	1.2	0.0	0.65	0.57
21	h32v09	11	-1.6	1.5	2.1	1.8	0.3	1.06	0.90
22	h32v10	7	1.2	1.0	1.5	1.4	-30.1	2.56	0.45
23	h33v09	322	-0.7	1.9	2.0	1.2	4.2	0.85	0.36
24	h33v10	117	-0.2	1.5	1.5	0.8	-0.7	1.04	0.67
25	h33v11	136	-0.1	1.3	1.3	0.6	1.2	0.92	0.44
Average		59145	0.0	1.4	1.5	1.0	2.7	0.77	0.51
Overall		59145	0.1	1.4	1.4	0.8	0.1	0.97	0.94

MTSAT. As such, a negative value indicates that MTSAT LST was found to be higher than the corresponding MODIS LST estimate. The overall daytime bias between the two LST products for 15 July 2007 was found to be -3.6 °C. This value based on the tile-by-tile comparison is much improved compared to the value for the full-mosaic comparison. Note that biases for individual tiles can reach as low as -8 °C (h30v10) and as high as 5.1 °C (h29v13), although the latter tile only had a negligible number of matching pixels with valid retrievals. The overall mean standard deviation and RMSE also have reasonable levels of 3.2 °C and 4.8 °C, respectively.

Table 13 shows the corresponding tile-by-tile results for nighttime data on 15 July 2007. As expected, the results are dramatically improved over the daytime statistics. The overall bias was found to be only 0.1 °C with the standard deviation and RMSE at very low values of 1.4 °C and 1.4 °C, respectively. The overall R² was also found to be very high with a value of 0.94.

Table 14 shows the results for the daytime comparison between MTSAT LST and MODIS LST for 15 December 2007. As previously, the biases are calculated as MODIS LST minus MTSAT LST, i.e. negative values indicate an overestimate of MTSAT LST with respect to the MODIS estimate. The overall bias was found to be quite high with a value of -4.9 °C, indicating again a systematic overestimation. The overall standard deviation between the two datasets is 5.4 °C, with an RMSE value of 7.3 °C. Note, however, that the total number of valid match-up pixels on this particular day was significantly lower than that for the 15 July 2007 dataset (6800 versus 60000), presumably due to increased cloud cover, which could explain some of the worse correspondence with the MODIS dataset.

Table 14 – Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for daytime data on 15 December 2007. Note that *Average* is just the mean of the results from each individual tile, i.e. there is no weighting by N, whereas *Overall* shows the statistics computed over the entire set of matching pixels independent of the respective tiles.

Tile No	Tile ID	N	Bias	StdDev	RMSE	MAD	Offset	Slope	R ²
1	h27v11	7	1.5	4.1	4.1	3.1	-10.0	1.23	0.22
2	h27v12	55	2.5	3.2	4.0	2.7	10.5	0.52	0.44
3	h28v09	0	0.0	0.0	0.0	0.0	0.0	0.00	0.00
4	h28v11	186	-2.0	4.6	5.0	3.5	4.1	0.95	0.59
5	h28v12	79	-2.6	2.2	3.4	3.2	4.3	0.95	0.55
6	h28v13	732	-1.8	3.5	3.9	2.3	3.7	0.90	0.65
7	h29v09	99	-2.0	2.4	3.1	2.5	7.4	0.82	0.32
8	h29v10	202	-10.4	5.2	11.6	10.4	15.1	0.87	0.47
9	h29v11	78	-7.0	4.1	8.1	7.3	17.5	0.79	0.55
10	h29v12	194	-1.1	4.0	4.1	3.3	-8.1	1.25	0.84
11	h29v13	73	10.2	3.9	10.9	10.1	12.7	-0.11	0.04
12	h30v09	6	-8.1	1.5	8.2	8.2	57.9	-0.88	0.15
13	h30v10	884	-12.3	2.1	12.5	12.4	12.9	0.98	0.94
14	h30v11	2206	-4.6	2.2	5.1	4.5	9.0	0.91	0.67
15	h30v12	500	-6.1	2.5	6.6	6.0	4.5	1.05	0.91
16	h31v09	0							
17	h31v10	310	-12.0	4.7	12.9	12.7	15.4	0.90	0.53
18	h31v11	718	0.7	4.4	4.5	2.9	9.4	0.71	0.34
19	h31v12	0							
20	h32v09	212	-4.2	3.1	5.3	4.9	14.5	0.60	0.12
21	h32v10	191	-4.9	2.8	5.6	4.9	-5.6	1.39	0.71
22	h33v09	34	0.0	2.1	2.1	1.6	12.9	0.53	0.06
23	h33v10	18	-2.6	1.4	2.9	2.3	-0.5	1.12	0.36
24	h33v11	13	-3.9	2.4	4.5	3.2	21.5	0.31	0.11
Average		6797	-3.2	3.0	5.8	5.1	9.5	0.72	0.43
Overall		6797	-4.9	5.4	7.3	4.8	3.7	1.03	0.82

Finally, Table 15 shows the corresponding tile-by-tile results for nighttime data on 15 December 2007. Again, the correspondence between the two data sources improves dramatically for nighttime retrievals. The overall bias for nighttime retrievals on 15 December 2007 was found to be only 0.2 °C, with a quite low standard deviation of only 1.9 °C. The RMSE as a measure of overall agreement was found to be equally low at 1.9 °C. Again, it should be noted that the overall number of pixel matchups between MODIS and MTSAT was significantly reduced for this day from 59145 to only 4452.

4.2.3 Dependence on zenith view angle

In addition to summary statistics the tile-by-tile comparison of MTSAT LST and MODIS LST also involved a brief investigation of the impact of the MODIS zenith view angle (ZVA) on the inter-comparison. Figure 37 show the scatterplot of daytime LST derived from both data sources for 15 July 2007, with an additional indication of the MODIS ZVA for each matchup. While the general daytime over-estimation of the MTSAT retrievals can be easily seen in the figure, the impact of the MODIS ZVA is not as clear. For most of the pixel matchups falling along the 1:1 line (particularly for those between 15 °C and 25 °C), the ZVA is relatively low on the order of less than 30 degrees. However, at very low temperatures of less than 10 °C, the MODIS ZVA was on the order of 50 degrees, yet the matchups are relatively close to the 1:1 lines (albeit slightly biased towards higher MTSAT LSTs). There is further a large cluster of matchup pixels beyond 20 °C (showing dark blue) for which MTSAT appears to severely overestimate with respect to

Table 15 – Tile-by-tile based summary statistics comparing MODIS MOD11 LST and MTSAT LST for nighttime data on 15 December 2007. Note that *Average* is just the mean of the results from each individual tile, i.e. there is no weighting by N, whereas *Overall* shows the statistics computed over the entire set of matching pixels independent of the respective tiles.

Tile No	Tile ID	N	Bias	StdDev	RMSE	MAD	Offset	Slope	R ²
1	h27v11	0							
2	h27v12	0							
3	h28v09	0							
4	h28v11	83	-0.7	1.3	1.5	1.2	-3.0	1.16	0.91
5	h28v12	62	1.9	1.1	2.2	1.8	-1.8	0.99	0.59
6	h28v13	43	-0.4	2.6	2.6	0.8	8.3	-0.08	0.01
7	h29v09	197	1.2	2.2	2.6	1.7	-7.8	1.28	0.44
8	h29v10	305	-1.7	2.4	2.9	1.6	-7.3	1.35	0.19
9	h29v11	107	-0.4	1.4	1.4	0.9	6.3	0.78	0.76
10	h29v12	41	-0.4	1.4	1.5	1.0	-4.4	1.35	0.97
11	h29v13	0							
12	h30v09	2	-3.5	0.8	3.6	3.5	-4.0	1.34	1.00
13	h30v10	39	-3.1	1.7	3.5	3.0	0.7	1.10	0.24
14	h30v11	183	-2.1	0.9	2.3	2.1	1.1	1.04	0.80
15	h30v12	25	0.2	1.6	1.5	1.0	-1.4	1.08	0.53
16	h31v09	0							
17	h31v10	1572	0.5	1.4	1.5	1.0	0.7	0.95	0.42
18	h31v11	1582	0.8	1.6	1.8	1.0	1.1	0.92	0.57
19	h31v12	78	0.4	1.2	1.3	0.8	4.9	0.65	0.39
20	h32v09	19	-1.0	1.7	1.9	1.8	-4.9	1.29	0.91
21	h32v10	79	-0.3	1.5	1.5	0.9	-2.7	1.13	0.86
22	h33v09	7	-2.9	1.6	3.2	2.8	-7.2	1.45	0.62
23	h33v10	6	-4.3	1.0	4.4	4.8	-2.5	1.29	0.79
24	h33v11	22	0.6	2.3	2.3	1.2	8.3	0.59	0.19
Average		4452	-0.8	1.6	2.3	1.7	-0.8	1.03	0.59
Overall		4452	0.2	1.9	1.9	1.1	-2.0	1.08	0.80

MODIS. This effect could be to some extent due to the very low ZVA of -50 degrees and more at which the MODIS LST was acquired here.

Figure 38 shows the corresponding comparison results for nighttime data. As mentioned previously, the nighttime data exhibit a much better correspondence between the two data sources with a nearly negligible bias. This is obvious from the figure as well, where the large majority of matchup pixels falls relatively close to the 1:1 line. Again, the majority of the matchup pixels that fall close to the 1:1 line have a ZVA of less than 30 degrees. Especially for lower temperatures of less than 10 °C the matchup pixels that are furthest away from the 1:1 line, i.e. show the highest biases between MODIS and MTSAT, exhibit positive MODIS ZVA values of more than 40 degrees. The exception are two small clusters of matchup pixels between 15 °C and 25 °C which also have relatively high positive MODIS ZVA, yet fall very close to the 1:1 line.

Figure 39 shows the MODIS versus MTSAT scatterplot of daytime LST for 15 December 2007. Similarly to Figure 37 it shows a large amount of scatter around the 1:1 line. Once again, the majority of matchup pixels close to the 1:1 line have relatively small absolute ZVA values of less than 30 degrees, but this is just a tendency and the relationship is not very strong.

Finally, Figure 40 shows the corresponding scatterplot for nighttime LST for 15 December 2007. Note that most matchup pixels have relatively small zenith view angles. Despite the significantly reduced number of matchups pixels this plot show the most promising relationship between MODIS LST and MTSAT LST. Both the bias as well as the scatter around the 1:1 line are very low and are comparable to what was shown for a comparison between AATSR LST and MODIS LST. This is quite remarkable as it indicates that for nighttime data an LST comparison between

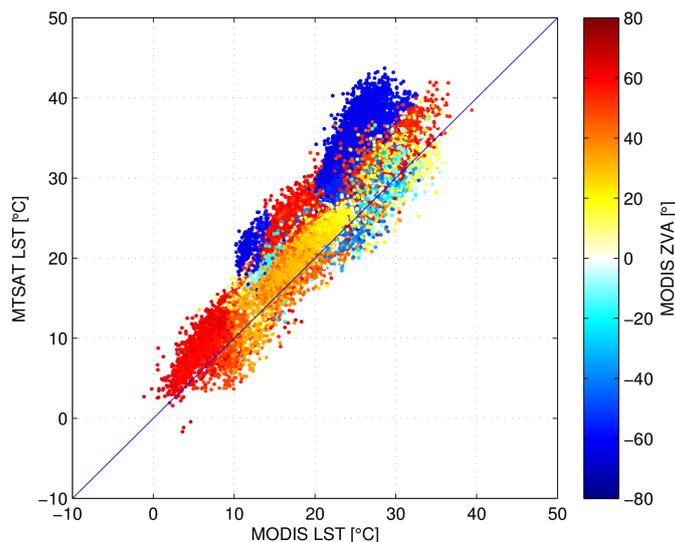


Figure 37 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 July 2007. MODIS-derived view zenith angle (ZVA) is also shown. Note that the matchups were derived from a tile-by-tile matching of MODIS overpass time and MTSAT observation time.

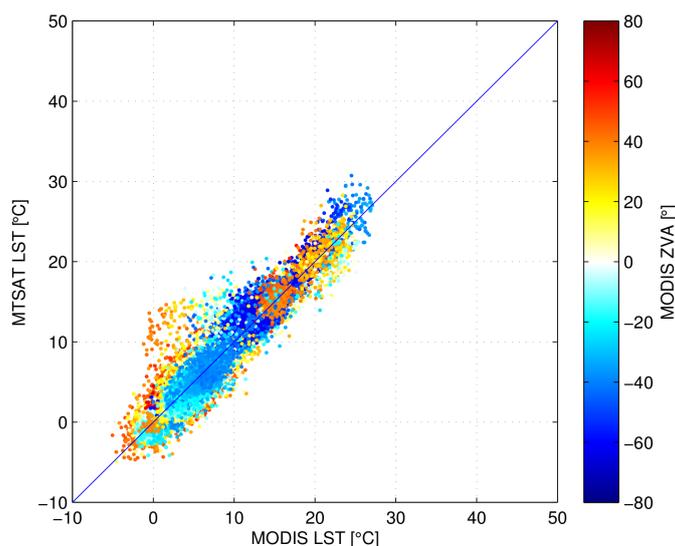


Figure 38 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 July 2007. MODIS-derived view zenith angle (ZVA) is also shown. Note that the matchups were derived from a tile-by-tile matching of MODIS overpass time and MTSAT observation time.

a geostationary platform (MTSAT) and a low-earth orbit platform (MODIS) can exhibit similarly low biases and standard deviation as the comparison between two low-earth orbit instruments.

In general it should be noted that, while it does exist, the relationship between inter-product biases and ZVA is quite weak. High MODIS zenith view angles are able to explain some of the large biases for some clusters of matchup pixels that are quite far from the 1:1 line, but large ZVA values alone cannot explain the substantial scattering daytime data.

Overall, the MTSAT LST product produced in the framework of WACMOS-ET corresponds quite well with MODIS LST for nighttime data, but it significantly over-estimates daytime LST with respect to MODIS. For nighttime data both the scatterplot and the error statistics derived for

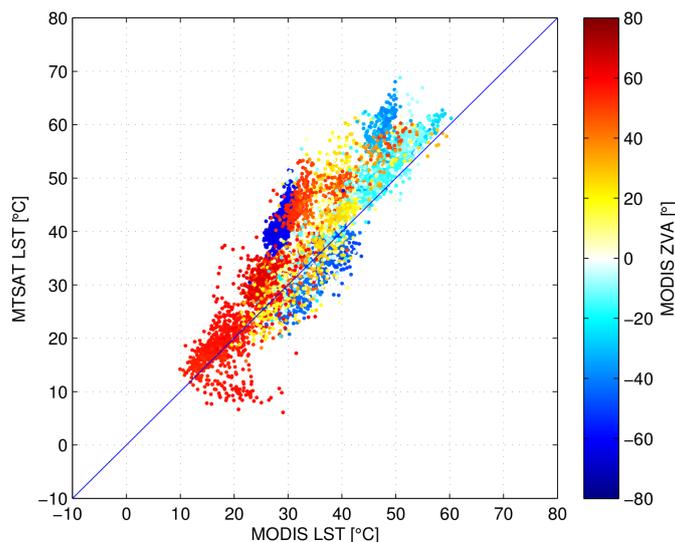


Figure 39 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for daytime data for 15 December 2007. MODIS-derived view zenith angle (ZVA) is also shown. Note that the matchups were derived from a tile-by-tile matching of MODIS overpass time and MTSAT observation time.

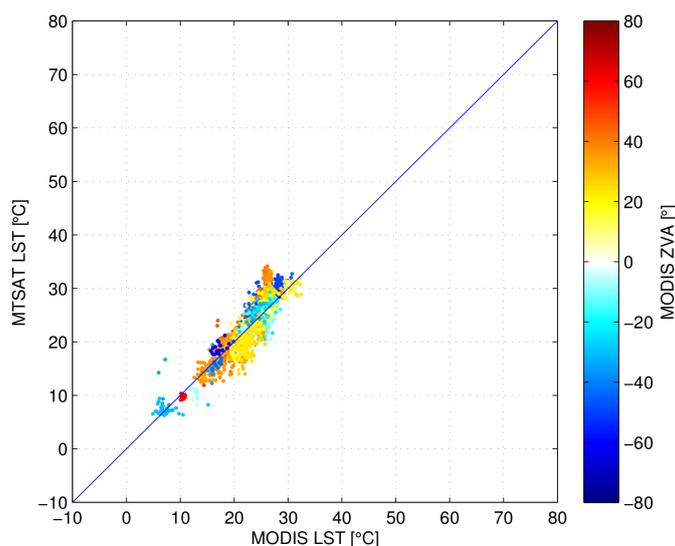


Figure 40 – Scatterplot showing the relationship between MODIS-based LST and MTSAT-derived LST for nighttime data for 15 December 2007. MODIS-derived view zenith angle (ZVA) is also shown. Note that the matchups were derived from a tile-by-tile matching of MODIS overpass time and MTSAT observation time.

the comparison between MTSAT LST and MODIS LST are along the lines of what would be expected and are similar to the results obtained for an inter-comparison between AATSR LST and MODIS LST (see Section 3.2.1). For daytime data, a significantly higher random error would be expected but the biases between the two products consistently reach around 5 °C, which is too much for most operational use. Unfortunately due to the lack of adequate in situ sites that are located in the geographical viewing extent of MTSAT no further quantitative conclusions can be drawn about the accuracy of the MTSAT LST product. Further work should attempt to locate adequate in situ stations in the view area of MTSAT to better evaluate the MTSAT LST product. Section 5 describing the validation of GOES-E LST shows that even non-dedicated in situ sites

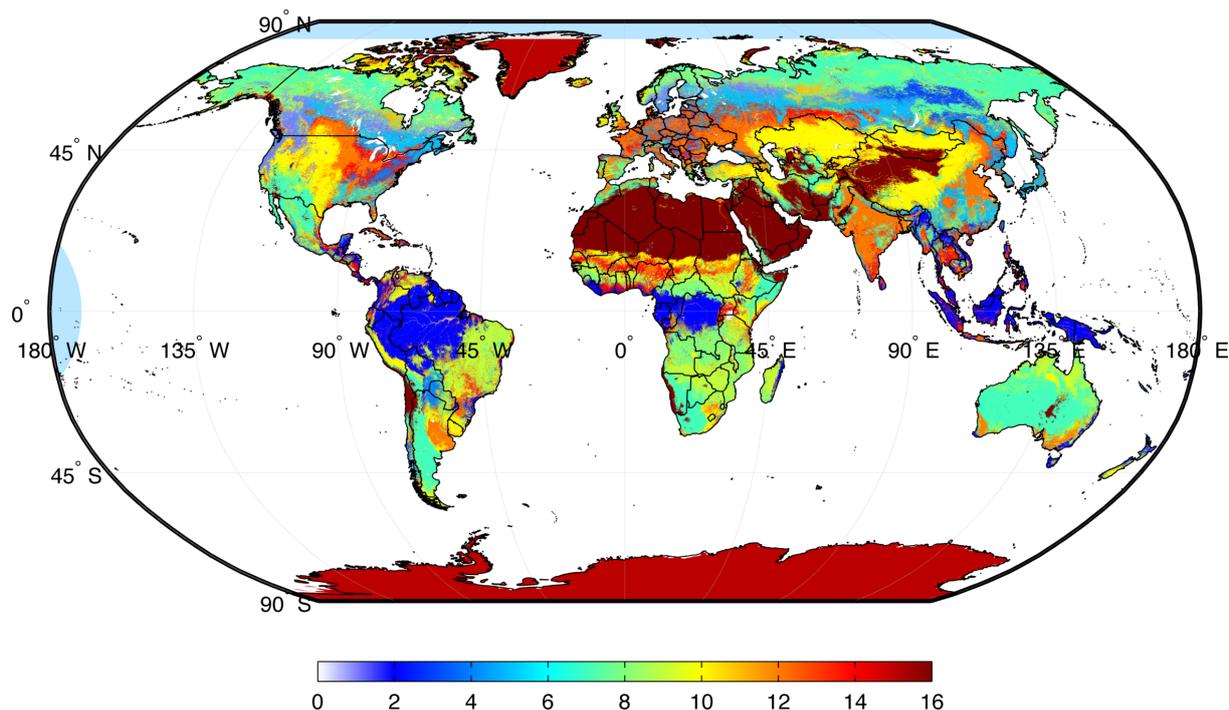


Figure 41 – Global map of land cover for 2007 as provided by the MODIS MCD12Q1 product. The IDs of the land cover classes are given in Table 16.

without land cover homogeneity can be used to validate LST from geostationary instruments and the comparison shows good performance with surprisingly small errors.

4.2.4 Dependence on land cover

The difference between MTSAT LST and MODIS LST was further investigated, based on the previous tile-by-tile inter-comparison, by analysing it with respect to land cover. Land cover was obtained from the MODIS MCD12Q1 product for 2007. This product provides an annual map of a land cover derived from the combined data of MODIS-Terra and MODIS-Aqua using multiple classification systems. For the purposes of this study, the classification proposed by the International Geosphere-Biosphere Programme (IGBP) was used. Figure 41 shows a global map of the spatial distribution of the IGBP land cover classes that were used in this study. The description of the land cover type for each class ID is provided in Table 16. Note that the following analysis was carried out using the full mosaics and not a tile-by-tile comparison.

Table 16 shows the relationship between MTSAT LST and MODIS LST characterized by land cover for daytime data only 15 July 2007. In both this and the following Tables the number of valid matchups (given as *N matchups* in the Table) should be kept in mind when assessing the biases and other error statistics. A very small number of matchups of less than 50 is unlikely to provide reliable statistics and those cases are ignored in the following description. As found in the previous sections for daytime data, the biases between the two instruments overall are quite high. Daytime MTSAT LST retrievals for 15 July 2007 were on average 5 °C higher than those by MODIS. The class with the highest absolute daytime bias is class 9 (Savannas), the class with the lowest bias is class 3 (Evergreen Broadleaf Forest), for which also the lowest standard deviation and RMSE, as well as the highest R^2 value were found.

Table 17 shows the land cover-characterized error statistics for the nighttime data on 15 July 2007. As seen before the nighttime data of the two instruments agree much more closely. The

Table 16 – Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for daytime data on 15 July 2007. Statistics were only computed for N matchups >50. All units except for N, slope, and R² in °C.

LC ID	Land cover	N gridcells	N matchups	Bias	StdDev	RMSE	MAD	Offset	Slope	R2	AvgTemp
0	Water	27517000	328	-2.51	3.44	4.25	2.13	2.17	1.02	0.73	19.97
1	Evergreen Needleleaf forest	153570	1	-1.60	0.00	1.60	1.60				
2	Evergreen Broadleaf forest	596450	1565	-2.24	2.07	3.05	2.40	1.88	1.03	0.93	24.36
3	Deciduous Needleleaf forest	65160	6	-9.29	5.26	10.46	11.82				
4	Deciduous Broadleaf forest	68407	37	-5.42	5.47	7.65	5.24				
5	Mixed forest	383740	158	-2.48	3.04	3.91	2.82	0.77	1.11	0.82	25.24
6	Closed shrublands	10550	199	-6.91	4.19	8.07	7.09	1.98	1.20	0.83	29.19
7	Open shrublands	947730	6077	-7.12	3.33	7.86	6.99	-0.17	1.34	0.83	31.62
8	Woody savannas	511040	14851	-5.51	2.81	6.18	5.50	5.16	1.02	0.76	25.86
9	Savannas	471920	19372	-9.09	4.12	9.98	7.66	14.05	0.81	0.34	28.02
10	Grasslands	876450	5507	-7.76	4.41	8.93	8.75	-3.38	1.49	0.89	33.63
11	Permanent wetlands	77708	293	-3.75	4.32	5.71	3.16	3.54	1.01	0.48	24.15
12	Croplands	554710	6467	-2.98	3.33	4.47	3.65	-2.78	1.38	0.83	30.01
13	Urban and built-up	30626	41	-3.87	3.80	5.39	3.85				
14	Cropland/Natural vegetation mosaic	422540	518	-5.84	6.23	8.54	4.53	4.19	1.07	0.43	28.46
15	Snow and ice	696560	1	0.25	0.00	0.25	0.25				
16	Barren or sparsely vegetated	887490	322	-7.13	3.11	7.78	7.89	5.94	1.06	0.47	43.82
	Average	2015979	3279	-4.90	3.47	6.12	5.02	2.78	1.13	0.69	28.69

Table 17 – Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for nighttime data on 15 July 2007. Statistics were only computed for N matchups >50. All units except for N, slope, and R² in °C.

LC ID	Land cover	N gridcells	N matchups	Bias	StdDev	RMSE	MAD	Offset	Slope	R2	AvgTemp
0	Water	27517000	82	0.16	3.10	3.08	1.73	-0.98	1.04	0.86	17.92
1	Evergreen Needleleaf forest	153570	0								
2	Evergreen Broadleaf forest	596450	3234	1.68	1.59	2.32	1.87	-2.25	1.09	0.95	18.64
3	Deciduous Needleleaf forest	65160	0								
4	Deciduous Broadleaf forest	68407	25	1.24	1.45	1.89	1.65				
5	Mixed forest	383740	252	1.44	1.84	2.34	1.81	-1.42	1.00	0.77	15.99
6	Closed shrublands	10550	198	0.72	1.41	1.58	0.96	-0.97	1.02	0.94	13.89
7	Open shrublands	947730	6433	0.87	1.48	1.71	1.19	-0.76	0.98	0.90	12.35
8	Woody savannas	511040	18094	1.04	1.38	1.73	1.25	-0.70	0.95	0.94	12.95
9	Savannas	471920	13797	1.38	1.41	1.97	1.57	-0.32	0.91	0.92	12.67
10	Grasslands	876450	4207	0.53	1.54	1.63	1.09	-0.70	1.03	0.93	13.16
11	Permanent wetlands	77708	114	0.36	2.25	2.26	1.65	-1.61	1.07	0.95	14.63
12	Croplands	554710	9224	0.79	1.15	1.40	0.93	-0.68	0.96	0.85	15.78
13	Urban and built-up	30626	61	0.88	1.34	1.60	0.98	-0.69	0.97	0.95	18.19
14	Cropland/Natural vegetation mosaic	422540	442	0.84	1.85	2.03	1.26	-0.59	0.97	0.95	16.51
15	Snow and ice	696560	3	3.40	1.35	3.58	3.05				
16	Barren or sparsely vegetated	887490	100	1.76	2.33	2.90	1.94	-0.48	0.83	0.66	22.53
	Average	2015979	3310	1.14	1.70	2.13	1.53	-0.93	0.99	0.89	15.78

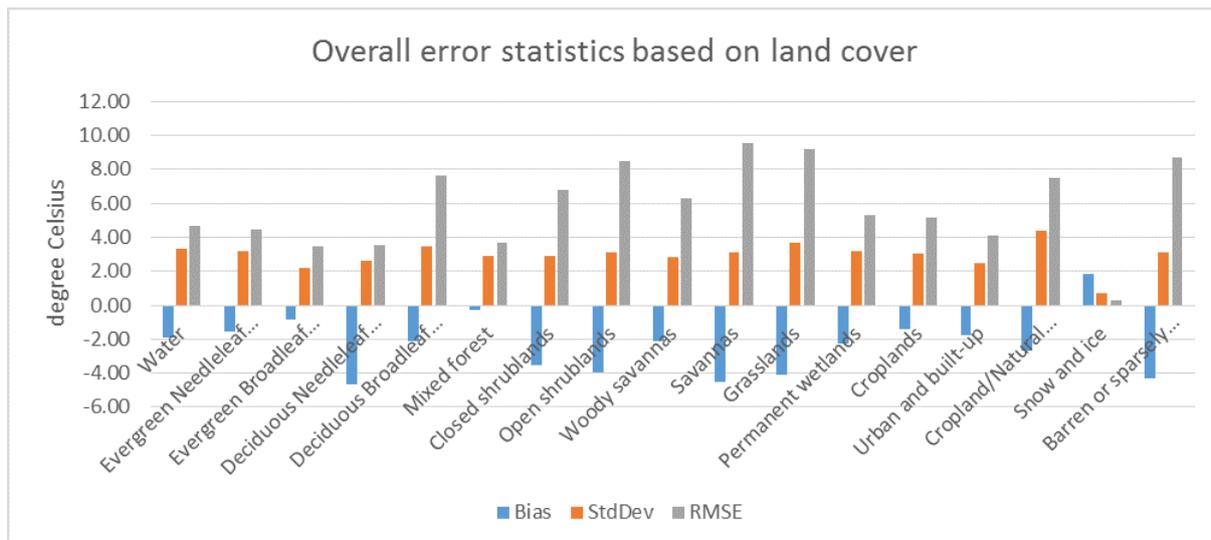


Figure 42 – Comparison of mean error statistics averaged for both daytime and nighttime data and the two test dates (15 July 2007 and 15 December 2007), shown by land cover class.

biases here are generally below 2 °C with the lowest values found for Water (class 0) and Permanent Wetlands (class 11). The lowest RMSE was found for Croplands (class 12) with a value of only 1.4 °C. Finally, the R^2 values for the majority of land cover classes were found to exceed 0.9.

Similarly Tables 18 and 19 show the corresponding results for 15 December 2007. Note for both Tables that the overall number of matchups between MTSAT and MODIS is significantly reduced over the July comparison. The daytime data shows again a very strong overestimation of MTSAT LST with respect to MODIS with the highest biases for Savannas (class 8 and 9) as well as Snow and Ice (class 15). The highest R^2 value was found for Closed shrublands (class 6).

Again, the comparison statistics for nighttime data are significantly improved. With an overall bias of only 0.47 °C, the lowest bias was found for Permanent Wetlands (class 11) with a value of -0.22 °C and the highest for Woody Savannas (class 8) with a value of 2.18 °C. The RMSE values are mostly in the 2-3 °C range, while the R^2 values are not as high as those observed for the July dataset but still reach values of 0.7 to 0.9.

Overall the analysis of MTSAT versus MODIS with respect to land cover is somewhat inconclusive. For daytime data, the biases and errors are so high across all classes that detailed conclusions are difficult. For nighttime data, the biases tend to be lowest for water surface and wetlands although there tend to be generally only few matchup pixels for both the July and the December test date.

Figure 42 shows the mean error statistics averaged over all four cases, i.e. nighttime and daytime and both test dates (15 July 2007 and 15 December 2007), by land cover class. It shows once again that the highest overall discrepancies as measured by the RMSE were found for savannas and grasslands, with open shrublands and barren areas following very closely. The lowest overall discrepancies were found for snow and ice surface, but it should be noted that the number of matchup pixels for this class was extremely low, so it is unlikely that these results are entirely representative. The inter-instrument bias was lowest for the Mixed forest class, with Evergreen broadleaf forest a close second.

Table 18 – Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for daytime data on 15 December 2007. Statistics were only computed for N matchups >50. All units except for N, slope, and R² in °C.

LC ID	Land cover	N gridcells	N matchups	Bias	StdDev	RMSE	MAD	Offset	Slope	R2	AvgTemp
0	Water	27517000	96	-4.09	4.03	5.73	4.24	7.35	0.88	0.55	5.09
1	Evergreen Needleleaf forest	153570	5	-1.58	6.36	5.90	5.79				
2	Evergreen Broadleaf forest	596450	668	-3.34	2.86	4.40	3.41	2.61	1.03	0.80	23.97
3	Deciduous Needleleaf forest	65160	1	-0.05	0.00	0.05	0.05				
4	Deciduous Broadleaf forest	68407	0								
5	Mixed forest	383740	16	-0.27	4.36	4.24	3.52				
6	Closed shrublands	10550	40	-7.52	4.24	8.61	6.89				
7	Open shrublands	947730	559	-10.54	5.73	11.99	10.24	7.95	1.06	0.65	19.37
8	Woody savannas	511040	552	-6.08	5.43	8.15	6.59	7.53	0.95	0.60	20.68
9	Savannas	471920	625	-11.89	4.97	12.88	11.86	21.25	0.70	0.31	31.45
10	Grasslands	876450	314	-10.71	6.55	12.55	11.45	4.07	1.19	0.78	12.94
11	Permanent wetlands	77708	53	-5.50	3.62	6.57	4.67	-1.09	1.24	0.63	1.55
12	Croplands	554710	144	-4.58	4.88	6.68	5.45	-7.90	1.40	0.78	18.93
13	Urban and built-up	30626	1	-3.66	0.00	3.66	3.66				
14	Cropland/Natural vegetation mosaic	422540	46	-6.45	6.85	9.35	6.06				
15	Snow and ice	696560	0								
16	Barren or sparsely vegetated	887490	963	-11.44	5.07	12.52	11.47	29.59	0.62	0.24	25.56
	Average	2015979	240	-5.85	4.33	7.55	6.36	7.93	1.01	0.59	17.73

Table 19 – Summary statistics for MODIS MOD11 LST minus MTSAT LST characterized by land cover class for nighttime data on 15 December 2007. Statistics were only computed for N matchups >50. All units except for N, slope, and R² in °C.

LC ID	Land cover	N gridcells	N matchups	Bias	StdDev	RMSE	MAD	Offset	Slope	R2	AvgTemp
0	Water	27517000	82	-1.17	2.60	2.84	1.61	-4.75	1.23	0.43	3.69
1	Evergreen Needleleaf forest	153570	0								
2	Evergreen Broadleaf forest	596450	253	0.44	2.11	2.15	1.42	0.31	0.96	0.87	18.60
3	Deciduous Needleleaf forest	65160	0								
4	Deciduous Broadleaf forest	68407	0								
5	Mixed forest	383740	18	0.24	2.28	2.23	1.85				
6	Closed shrublands	10550	57	-0.52	1.87	1.92	1.29	-5.95	1.25	0.22	11.14
7	Open shrublands	947730	423	0.96	1.92	2.15	1.35	-5.38	1.18	0.74	4.49
8	Woody savannas	511040	1625	2.18	1.58	2.69	2.14	-3.29	1.05	0.67	10.80
9	Savannas	471920	630	1.52	2.07	2.57	1.81	-10.03	1.36	0.59	17.69
10	Grasslands	876450	356	1.66	2.23	2.78	2.11	-0.88	0.97	0.69	-5.92
11	Permanent wetlands	77708	80	-0.22	2.46	2.45	1.67	8.54	0.67	0.19	-3.02
12	Croplands	554710	66	1.14	2.74	2.95	2.10	-4.90	1.20	0.90	3.50
13	Urban and built-up	30626	3	-0.48	4.63	3.81	2.97				
14	Cropland/Natural vegetation mosaic	422540	112	0.75	2.56	2.66	1.71	-0.80	1.00	0.50	9.58
15	Snow and ice	696560	0								
16	Barren or sparsely vegetated	887490	132	-0.39	1.94	1.97	1.19	4.22	0.86	0.59	5.04
	Average	2015979	226	0.47	2.38	2.55	1.79	-2.08	1.07	0.58	6.87

5 Validation of GOES

An LST product derived from GOES-E data was also produced within the framework of the WACMOS-ET project. The validation of this product was primarily carried out using a comparison against in situ data at SURFRAD stations in the United States.

5.1 Validation against in situ data

Similarly to the comparison against in situ data that was carried out for the validation of AATSR data, GOES-E LST was also validated against observation acquired at the SURFRAD stations in the United States.

Figure 43 shows scatterplots for 7 SURFRAD stations for daytime data. Qualitatively it can be said that for nearly all stations the data points follow the 1:1 line remarkably well, considering the vast differences in spatial representativeness between the GOES-E footprint and the in situ observations and the highly heterogeneous land cover and emissivity at most sites. The Bondville station indicates slightly more scatter for LSTs of more than 25 °C. A similar effect can be seen at Sioux Falls although it is much weaker. A very good relationship between the two data sources can be observed at the Desert Rock station which indicates a mostly unbiased relationship all the way from 0 °C to nearly 60 °C, and with only a comparatively small amount of

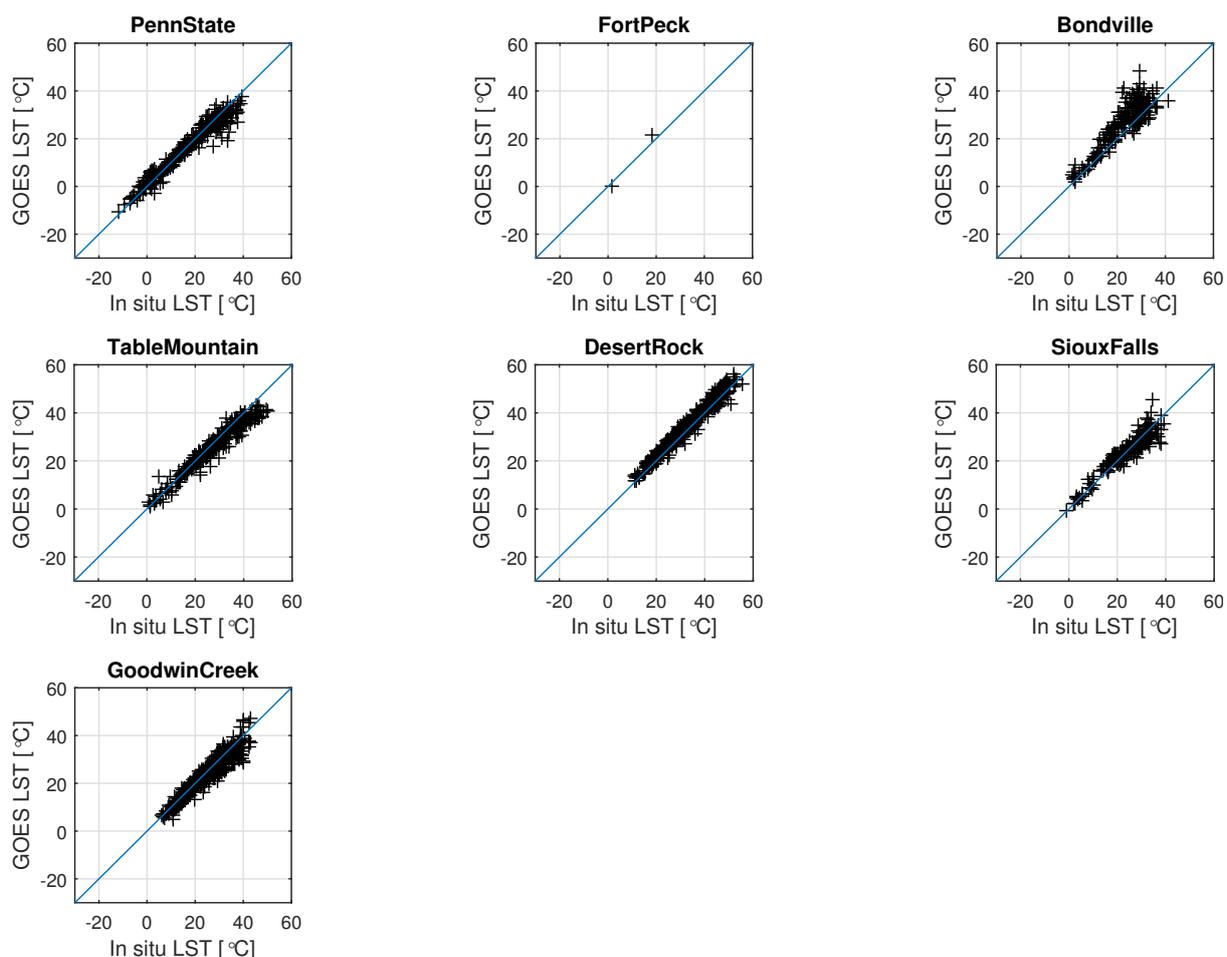


Figure 43 – Scatter plots of original GOES-derived daytime LST against LST computed from observations at in situ stations.

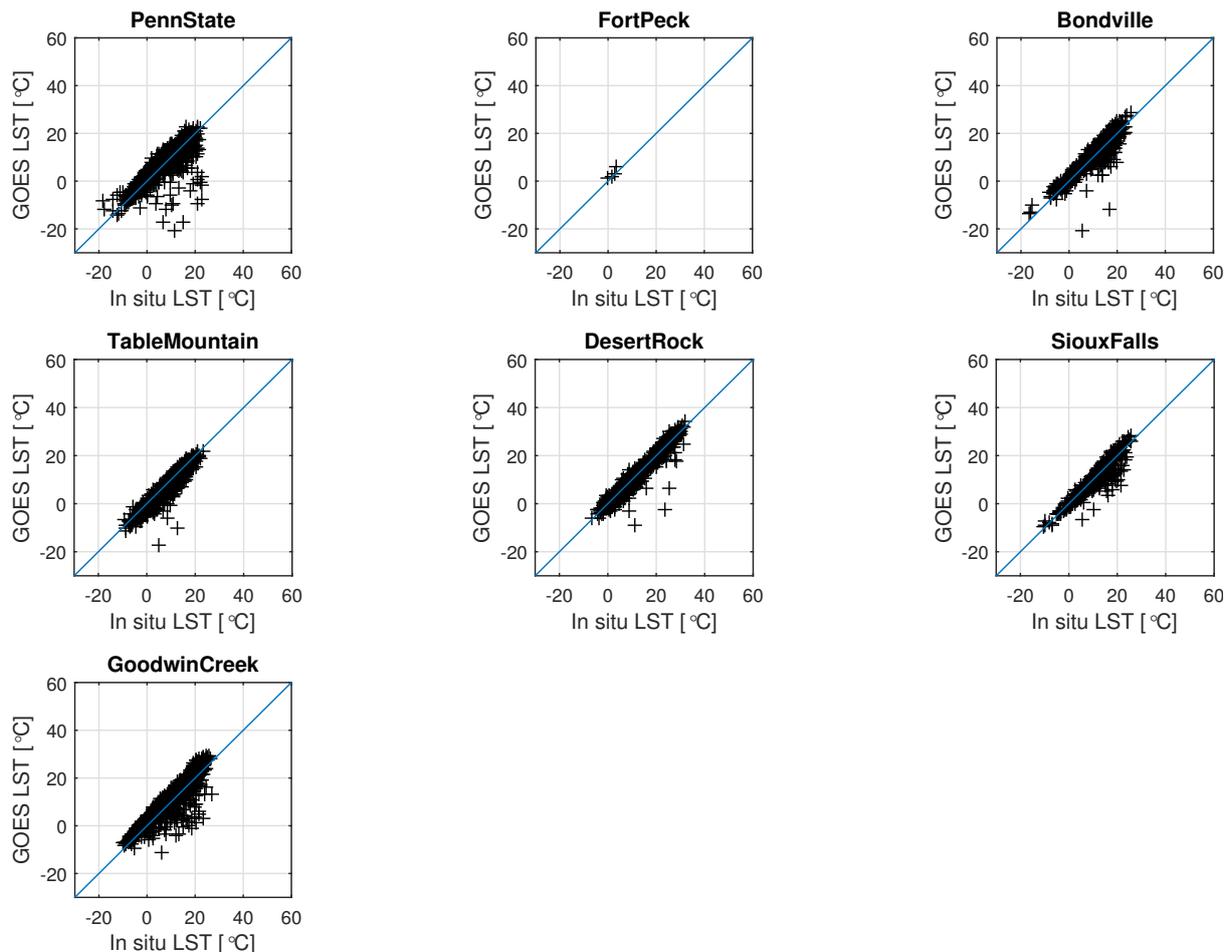


Figure 44 – Scatter plots of original GOES-derived nighttime LST against LST computed from observations at in situ stations.

scatter (given that this is daytime data). The Fort Peck station only shows a very small number of matchups between GOES-E and the in situ data. The reason is not entirely clear and will need some further investigation into the raw datasets.

Figure 44 shows the scatterplots for nighttime data. While the majority of matchups are along the 1:1 line, a clear issue with negative outliers can be observed, particularly at the Penn State and Goodwin Creek stations but also to some extent at the other stations. Similarly to the cloud masking issues observed for AATSR, these outliers are due to the cloud masking procedure not having access to the visible channel at night and thus not performing as well as during daytime.

As was done previously for AATSR, a very simple correction scheme to eliminate erroneous outliers due to cloud masking problems was applied. This involved the removal of outliers which showed an difference to the in situ data of more than three standard deviations. While this does not affect the daytime data to a large extent, the cloudy matchups in the nighttime data are mostly removed and the derived statistics are then much more representative of what can be achieved with the actual algorithm. More research in improving the nighttime cloud mask will be necessary in future.

Figure 45 and 46 show the scatterplots for the corrected data after the statistical cloud removal scheme. The daytime scatterplots are mostly unchanged and for a quantitative comparison, Table 20 shows the corresponding statistics for corrected daytime data. The average bias was found to be only 0.37 °C, with the minimum absolute value exhibited by the Sioux Falls station

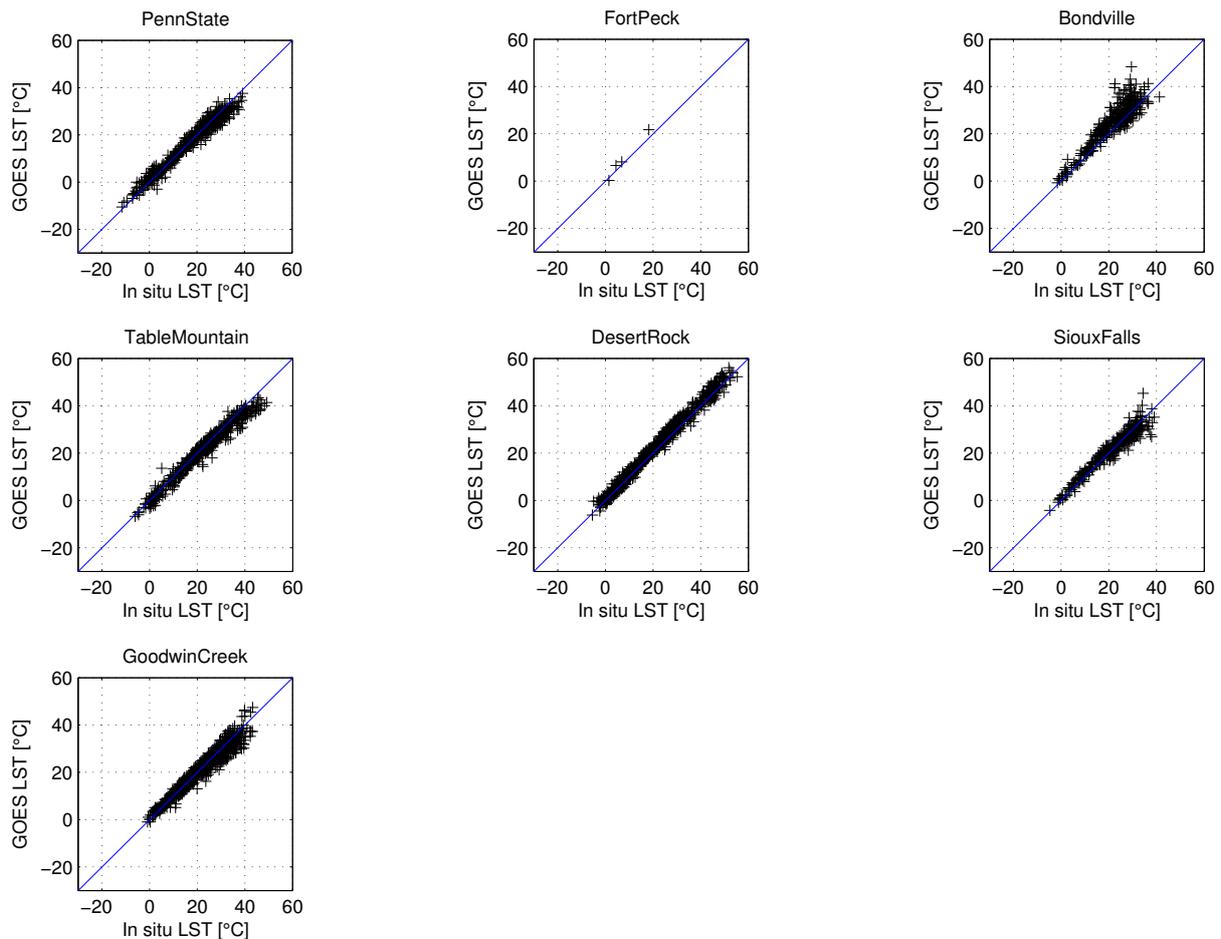


Figure 45 – Scatter plots of GOES-derived daytime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.

(-0.28 °C) and the maximum by the Bondville station (3.45 °C). The standard deviations have an average of 2.94 °C, where again the Bondville station performs worst with 4.34 °C and Desert Rock performs best with only 1.82 °C. The RMSE as a measure of overall accuracy was found to be 3.18 °C on average.

For nighttime data the correction scheme helped significantly and Figure 46 shows that qualitatively the results now look very promising as the data points tend to generally follow the 1:1 lines quite well. Again, the Bondville and Sioux Falls stations show some slight irregularities including a few negative outliers at lower temperatures. Table 21 shows the corresponding statistics for the nighttime comparison. As expected, the nighttime data from GOES-E matches the in situ dataset even better with an average bias of only 0.21 °C. The highest absolute bias was observed at the Table Mountain station with -1.26 °C, whereas the lowest absolute bias was found for the Desert Rock station with only -0.15 °C. The standard deviations are also reduced with respect to the daytime comparison and now reach an average value of only 2.4 °C, with the maximum of 3.42 °C found at the Penn State station and the minimum with 1.66 °C at Desert Rock (not counting the For Peck station which only had a very low number of matchups of N=4). The average RMSE was found to be 2.56 °C with a maximum of 3.48 °C at the Penn State station and a minimum of 1.67 °C at the Desert Rock station.

Finally, Figure 47 shows the scatterplot and the summary statistics for the overall comparison between GOES-E LST at all sites for both daytime and nighttime. The overall scatter looks quite reasonable, following the 1:1 line relatively closely without any major discernible bias. The vast

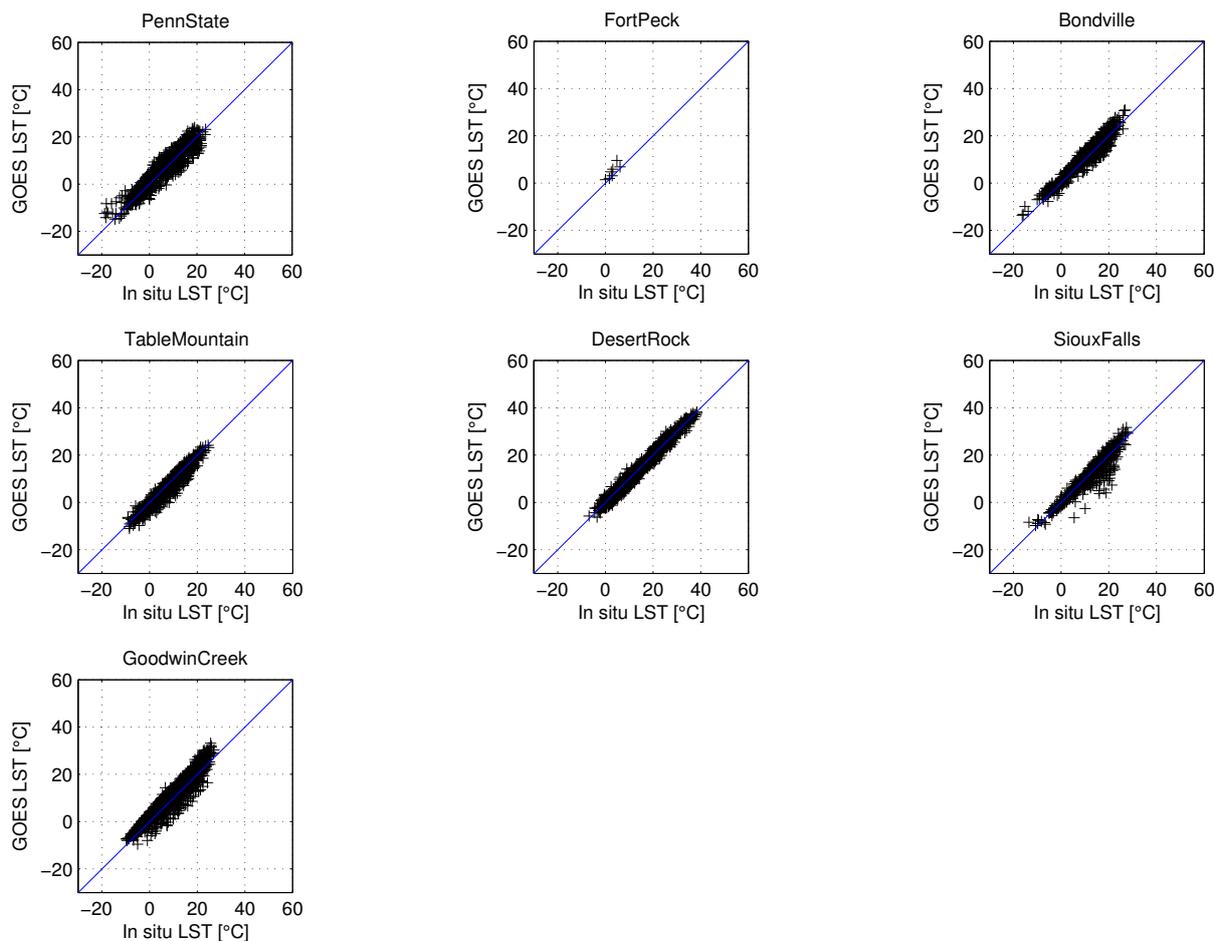


Figure 46 – Scatter plots of GOES-derived nighttime LST against LST computed from observations at in situ stations after additional statistical cloud filtering.

Table 20 – Summary statistics of the validation results for corrected GOES-derived daytime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Penn State	288	-0.54	2.74	2.78
Fort Peck	2	1.14	3.32	2.61
Bondville	185	3.45	4.34	5.54
Table Mountain	252	-1.74	2.81	3.30
Desert Rock	346	1.49	1.82	2.35
Sioux Falls	177	-0.28	3.00	3.01
Goodwin Creek	472	-0.91	2.52	2.68
Average	246	0.37	2.94	3.18

majority of matchup pixels was obtained for temperatures between -10 °C and +40 °C but even some outliers down to -20 °C and up to +60 °C do not show a significantly increases amount of scatter. Some of the very low temperature observations from GOES-E show a slight bias towards higher values as compared to the in situ data. The statistics indicate a very low bias overall of only 0.08 °C with a standard deviation of 3.5 °C and an RMSE of also 3.5 °C. A fitted linear regression model shows a very low offset of 0.5 °C with a slope of very close to unity (0.97) and an R² value of 0.9.

Table 21 – Summary statistics of the validation results for corrected GOES-derived nighttime LST against LST computed from observations at in situ stations. All values except the number of matchups N are given in units of °C.

Station	N	Bias	Std Dev	RMSE
Penn State	604	0.64	3.42	3.48
Fort Peck	4	1.20	1.18	1.58
Bondville	525	0.68	2.43	2.52
Table Mountain	591	-1.26	2.11	2.46
Desert Rock	926	-0.15	1.66	1.67
Sioux Falls	394	-0.65	2.81	2.88
Goodwin Creek	875	1.00	3.19	3.34
Average	560	0.21	2.40	2.56

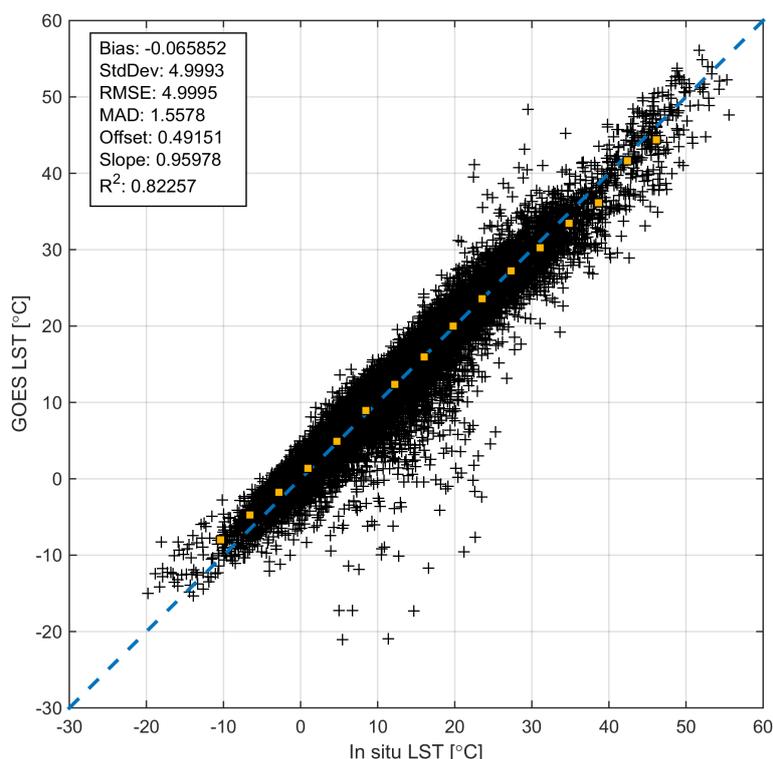


Figure 47 – Overall scatterplot including summary statistics of all matchups between cloud-corrected GOES LST and in situ LST at all stations and all times. Yellow markers indicate the median GOES LST for various classes of in situ LST.

Overall, these results agree with those reported in a similar study in which GOES-8 imager data was validated using the SURFRAD stations (Yu et al., 2012).

5.2 Validation against independent EO data

A very simple inter-comparison was carried out between the LST for GOES-E and the MODIS MOD11 product. When preparing a continental scale mosaic of MODIS tiles to compare against a geostationary instrument, the same issue arises that was found for a similar analysis for MTSAT: It is not possible to produce a mosaic from low-earth orbit sensors that has the same observation time as the geostationary sensor as that latter sees the entire coverage area at once. To make matters worse, even when a relatively close (in time) matchup is available for one area of the continental-scale map, other areas will have much larger difference in respective observation

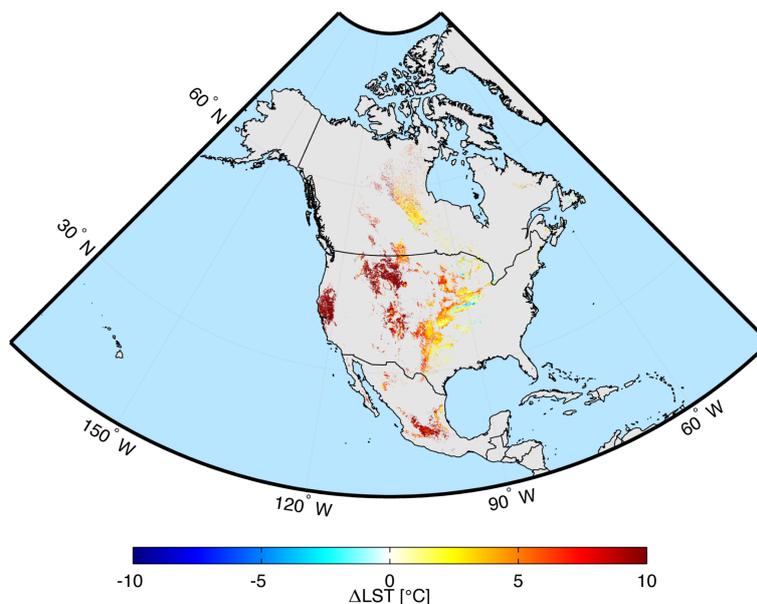


Figure 48 – Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and GOES-derived LST for daytime data for 15 July 2007. The daytime MODIS mosaic was matched with GOES-E data observed at 14:45 UTC.

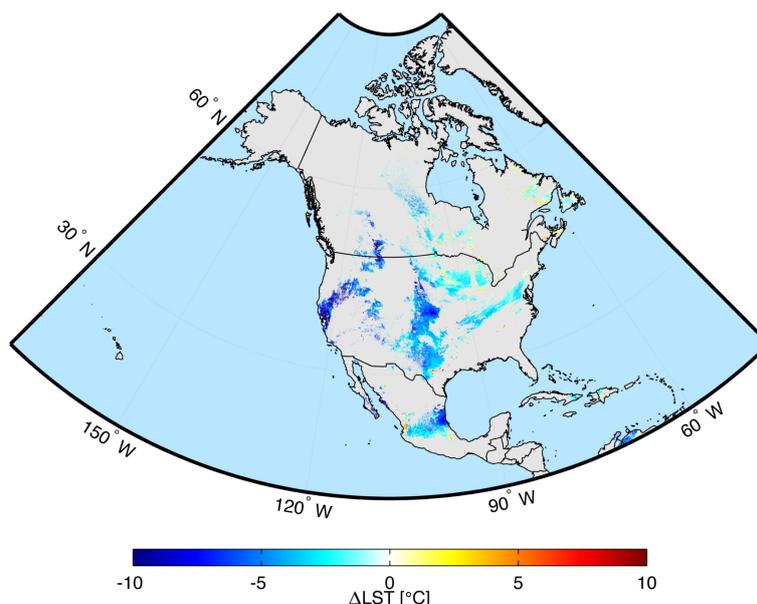


Figure 49 – Difference image showing the spatial patterns of the discrepancies between MODIS-based LST and GOES-derived LST for nighttime data for 15 July 2007. The nighttime MODIS mosaic was matched with GOES-E data observed at 02:45 UTC.

times. While an approach such as this is possibly feasible for nighttime data when even relatively large differences in observation time do not have a very large impact on the temperatures, for daytime data the difference images essentially only show a longitudinal effect of observation time differences, rather than true differences in LST between the sensors.

This effect can be clearly seen in Figures 48 and 49. A matchup time was chosen for MODIS to most reflect the conditions in the Eastern United States, where most of the SURFRAD in situ sites are located. The daytime map in general has mostly positive values, indicating that MODIS provided higher values than GOES-E. The opposite is true for the nighttime map, which shows that GOES-E had higher LST values than MODIS MOD11. In general, these figures should be

interpreted with caution, in particular the one for daytime data, as they show mostly the effect of difference in observation time rather than true LST differences.

While in general inter-comparisons between LST products from different sensors can be valuable, this is mostly helpful for comparisons between two low-earth orbit instruments with roughly similar overpass times. Inter-comparisons between low-earth orbit instruments and geostationary instruments are much more challenging and more difficult to interpret due to the large differences in observation which can occur when investigating relatively large areas. In some cases when absolutely no usable in situ dataset is available, as it was the case for the MTSAT validation (Section 4) reported on earlier here, it can be still be helpful to perform such a comparison to get a general idea about the product performance. However, when in situ information is available, such as in the case of GOES-E, it is strongly recommended to rely on a direct validation against the in situ observations rather than on the instrument inter-comparison. For this reason, and since the direct validation against in situ observations showed very good results for GOES-E, no further inter-comparison analysis was carried out for GOES-E.

6 Validation of SEVIRI

While LST data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) is being used within the framework of the WACMOS-ET project, the accuracy of the operational Land-SAF SEVIRI LST product was not evaluated as it has been validated extensively in the past and is known to provide LST estimates which are in good agreement with in situ observations, with RMS errors ranging between 1 °C and 2 °C (Kabsch et al., 2008; Trigo et al., 2008a; Freitas et al., 2010; Göttsche et al., 2013). Figure 50 shows an example of the quality of MSG/SEVIRI data as validated at the Gobabeb station in Namibia by Göttsche et al. (2013).

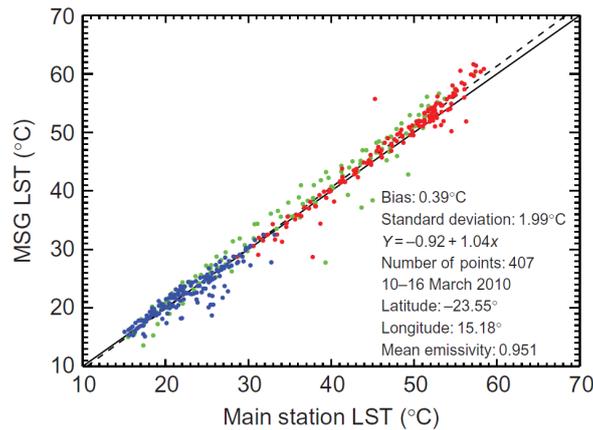


Figure 50 – LST from the Gobabeb station versus MSG/SEVIRI LST from the Land-SAF for 10-16 March 2010 (from Göttsche et al. (2013)).

7 Conclusions

The suite of WACMOS-ET LST products derived from AATSR, MTSAT, and GOES data has been evaluated using extensive comparisons with in situ reference datasets as well as inter-comparisons with other well-validated satellite-based LST products. This methodology therefore follows two of the main validation categories recommended in a recent white paper on LST validation (Schneider et al., 2012). The results obviously vary from instrument to instrument, but overall the LST products generated within the framework of WACMOS-ET appear to have a quite reasonable quality. The validation presented here is considered to be somewhat representative of the LST performance under real-world conditions (i.e. does not only provide statistics at dedicated LST validation sites with extremely homogeneous surroundings, which are unlikely to occur in areas where satellite-based LST products tend to be generally used). As expected, for all instruments the nighttime LST retrievals far outperform the daytime retrievals and it is recommended to use nighttime data whenever possible.

AATSR LST showed a very good relationship with ground-based in situ observations. One initial issue was found to be the AATSR cloud mask which did not perform in a sufficient manner, particularly for nighttime data when the visible channels were not available to provide additional information for clouds masking. After statistically filtering out likely outliers due to undetected clouds, the AATSR LST matched the in situ LST quite well. Daytime AATSR retrievals exhibited a bias of 1.4 °C on average with a standard deviation of 2.5 °C and an RMSE of 3.2 °C. The nighttime data showed a very low average bias of only 0.3 °C with a standard deviation of 1.3 °C and an overall RMSE of 1.4 °C. It should be noted that despite the cloud masking issues these values indicate that the WACMOS-ET AATSR LST product outperforms the MODIS-Terra MOD11A1 as the latter had some significant issues with negative biases which have also been reported by other studies previously (Wang et al., 2008). Regarding the cloud masking issues it might be worthwhile to explore other cloud masking techniques such as those proposed by Merchant et al. (2005) and Bulgin et al. (2014). Furthermore it was found that the WACMOS-ET AATSR LST product provides slightly higher accuracies than the GlobTemperature AATSR product when the same cloud mask is used for both products.

LST from MTSAT was extremely challenging to evaluate due to a number of factors. There was no usable dedicated in situ site measuring downwelling and upwelling thermal infrared radiation in the field of view of MTSAT. One site from the ARM network in Darwin, Australia, was investigated to some extent, but a true validation could not be carried out there as the corresponding MTSAT pixel would include a large fraction of ocean. In addition, the site is in a very complex urban landscape and not well suited for LST validation. As a validation against in situ data was not possible for MTSAT LST, the main focus shifted on inter-comparing the data with independent data from MODIS. The quantitative results from a tile-by-tile-based comparison showed a bias of -3.6 °C for daytime data (indicating that MTSAT overestimates LST with respect to MODIS) and 0.1 °C for nighttime data. The standard deviations were found to be 3.2 °C and 1.4 °C, respectively. It should be noted that all the inter-comparisons between MTSAT and MODIS were complicated by sometimes significant differences in observation time as well as the fact that the MODIS LSTs were shown to be biased low by 1-2 °C in the in situ comparison. No strong relationship between inter-sensor biases and zenith view angle were found. An investigation of the dependence of inter-instrument biases on land cover revealed no strong relationship although water surfaces and wetlands tended to have the lowest biases (albeit for a relatively small sample size).

Finally, LST from GOES-E was evaluated against in situ observations from SURFRAD stations. The results indicate a very good correspondence with an average bias of 0.21 °C and a standard deviation of 2.4 °C. The RMSE as an overall representation of error was 2.56 °C. A simple

inter-comparison with MODIS MOD11 data was carried out but the difference in observation times were too large to draw any quantitative conclusions.

Overall, the suite of LST products derived within the framework of the WACMOS-ET project are shown to have a good quality. Particularly AATSR shows a very good overall performance, aside from issues with the nighttime cloudmask. GOES-E shows surprisingly good correlations with the in situ data for being a geostationary instrument with only a single thermal infrared channel that was furthermore evaluated at non-dedicated validation sites. MTSAT LST could not adequately be validated against in situ LST observations, but an inter-comparison with independent remote sensing data showed reasonable agreement for nighttime data.

Bibliography

- Augustine, J. A., DeLuisi, J. J., and Long, C. N. (2000). SURFRAD - A National Surface Radiation Budget Network for Atmospheric Research. *Bulletin of the American Meteorological Society*, 81(10):2341–2357.
- Augustine, J. A., Hodges, G. B., Cornwall, C. R., Michaelsky, J. J., and Medina, C. I. (2005). An Update on SURFRAD - The GCOS Surface Radiation Budget Network for the Continental United States. *Journal of Atmospheric and Oceanic Technology*, 22:1460–1472.
- Bulgin, C., Sembhi, H., Ghent, D., Remedios, J., and Merchant, C. (2014). Cloud-clearing techniques over land for land-surface temperature retrieval from the Advanced Along-Track Scanning Radiometer. *International Journal of Remote Sensing*, 35(10):3594–3615.
- Coll, C., Caselles, V., Galve, J., Valor, E., Niclos, R., Sanchez, J., and Rivas, R. (2005). Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data. *Remote Sensing of Environment*, 97(3):288–300.
- Coll, C., Wan, Z., and Galve, J. M. (2009). Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product. *Journal of Geophysical Research*, 114(D20):1–15.
- Ermida, S. L., Trigo, I. F., DaCamara, C. C., Göttsche, F. M., Olesen, F. S., and Hulley, G. (2014). Validation of remotely sensed surface temperature over an oak woodland landscape - The problem of viewing and illumination geometries. *Remote Sensing of Environment*, 148:16–27.
- Freitas, S. C., Trigo, I. F., Bioucas-Dias, J. M., and Göttsche, F.-M. (2010). Quantifying the Uncertainty of Land Surface Temperature Retrievals From SEVIRI / Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):523–534.
- Galve, J. M., Coll, C., Caselles, V., Valor, E., and Nicl, R. (2007). Simulation and validation of land surface temperature algorithms for MODIS and AATSR data. *Tethys, Journal of Weather and Climate of the Western Mediterranean*, 4(1975):27–32.
- Göttsche, F., Olesen, F., and Bork-Unkelbach, A. (2011). Validation of Operational Land Surface Temperature Products with Three Years of Continuous In-Situ Measurements. In *Proceedings of the 2011 EUMETSAT Meteorological Satellite Conference*, Oslo, Norway, 5-9 September 2011.
- Göttsche, F.-M., Olesen, F.-S., and Bork-Unkelbach, A. (2013). Validation of land surface temperature derived from MSG/SEVIRI with in situ measurements at Gobabeb, Namibia. *International Journal of Remote Sensing*, 34(9-10):3069–3083.
- Guillevic, P., Privette, J., and Coudert, B. (2012). Land Surface Temperature product validation using NOAA's surface climate observation networks - Scaling methodology for the Visible Infrared Imager Radiometer. *Remote Sensing of Environment*, 124:282–298.
- Guillevic, P. C., Biard, J. C., Hulley, G. C., Privette, J. L., Hook, S. J., Olioso, A., Göttsche, F. M., Radocinski, R., Román, M. O., Yu, Y., and Csiszar, I. (2014). Validation of Land Surface Temperature products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements. *Remote Sensing of Environment*, 154:19–37.
- Guillevic, P. C., Bork-unkelbach, A., Göttsche, F. M., Hulley, G., Olesen, F. S., and Privette, J. L. (2013). Directional Viewing Effects on Satellite Land Surface Temperature Products Over

- Sparse Vegetation Canopies - A Multisensor Analysis. *IEEE Geoscience and Remote Sensing Letters*, 10(6):1464–1468.
- Hook, S. J., Prata, F. J., Alley, R. E., Abtahi, A., Richards, R. C., Schladow, S. G., and Pálmarsson, S. Ó. (2003). Retrieval of Lake Bulk and Skin Temperatures Using Along-Track Scanning Radiometer (ATSR-2) Data: A Case Study Using Lake Tahoe, California. *Journal of Atmospheric and Oceanic Technology*, 20(4):534.
- Hook, S. J., Vaughan, R. G., Tonooka, H., and Schladow, S. G. (2007). Absolute Radiometric In-Flight Validation of Mid Infrared and Thermal Infrared Data From ASTER and MODIS on the Terra Spacecraft Using the Lake Tahoe, CA/NV, USA, Automated Validation Site. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1798–1807.
- Hulley, G., Veraverbeke, S., and Hook, S. (2014). Thermal-based techniques for land cover change detection using a new dynamic MODIS multispectral emissivity product (MOD21). *Remote Sensing of Environment*, 140:755–765.
- Kabsch, E., Olesen, F. S., and Prata, F. (2008). Initial results of the land surface temperature (LST) validation with the Evora, Portugal ground truth station measurements. *International Journal of Remote Sensing*, 29(17-18):5329–5345.
- Kalnay, E., Kanamitsu, M., Kistler, R., and Collins, W. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3):437–471.
- Martins, J. P., Trigo, I. F., and Pires, A. C. (2014). WACMOS-ET LST Product Algorithm Theoretical Basis Document. Technical report, Instituto Português do Mar e da Atmosfera, Lisbon, Portugal.
- Merchant, C. J., Harris, A. R., Maturi, E., and Maccallum, S. (2005). Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quarterly Journal of the Royal Meteorological Society*, 131(611):2735–2755.
- Ogawa, K., Schmugge, T., and Jacob, F. (2003). Estimation of land surface window (8-12 μm) emissivity from multi-spectral thermal infrared remote sensing - A case study in a part of Sahara Desert. *Geophysical Research Letters*, 30(2):1067.
- Olesen, F.-S. and Göttsche, F.-M. (2009). Validation of Land Surface Temperatures Obtained from METEOSAT-MVIRI and SEVIRI with in-situ measurements. In *2009 EUMETSAT Meteorological Satellite Conference, Bath, UK, 21-25 September 2009*. EUMETSAT.
- Prata, A. J. (1994a). Land surface temperatures derived from the advanced very high resolution radiometer and the along-track scanning radiometer: 2. Experimental results and validation of AVHRR algorithms. *Journal of Geophysical Research*, 99(D6):13025–13058.
- Prata, A. J. (1994b). Validation Data for Land Surface Temperature Determination from Satellites. Technical Report 33, CSIRO, Division of Atmospheric Research, Aspendale, Australia.
- Prata, F. (2002). Land surface temperature measurement from space: AATSR algorithm theoretical basis document. Technical report, CSIRO Division of Atmospheric Research, Aspendale, Australia.
- Prata, F. (2003). Land surface temperature measurement from space: Validation of the AATSR Land Surface Temperature Product. Technical report, CSIRO Division of Atmospheric Research, Aspendale, Australia.
- Renzullo, L. J. (2009). MTSAT-1R land surface temperature retrievals and their potential role as observational constraints on surface energy balance for Australia. Technical report, CSIRO.

- Schneider, P., Ghent, D., Prata, F., Corlett, G. K., and Remedios, J. J. (2012). AATSR Validation: LST Validation Protocol. Technical report, NILU - Norwegian Institute for Air Research and University of Leicester.
- Schneider, P. and Hook, S. J. (2010). Space observations of inland water bodies show rapid surface warming since 1985. *Geophysical Research Letters*, 37(22):1–5.
- Schneider, P., Hook, S. J., Radocinski, R. G., Corlett, G. K., Hulley, G. C., Schladow, S. G., and Steissberg, T. E. (2009). Satellite observations indicate rapid warming trend for lakes in California and Nevada. *Geophysical Research Letters*, 36(22):1–6.
- Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L. (2008). Development of a Global Infrared Land Surface Emissivity Database for Application to Clear Sky Sounding Retrievals from Multispectral Satellite Radiance Measurements. *Journal of Applied Meteorology and Climatology*, 47(1):108–123.
- Stokes, G. M. and Schwartz, S. E. (1994). The Atmospheric Radiation Measurement (ARM) Program - Programmatic Background and Design of the Cloud and Radiation Test Bed. *Bulletin of the American Meteorological Society*, 75(7):1201–1221.
- Trigo, I. F., Monteiro, I. T., Olesen, F., and Kabsch, E. (2008a). An assessment of remotely sensed land surface temperature. *Journal of Geophysical Research*, 113(D17):1–12.
- Trigo, I. F., Peres, L. F., Dacamara, C. C., and Freitas, S. C. (2008b). Thermal Land Surface Emissivity Retrieved From SEVIRI / Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2):307–315.
- Wan, Z. (2007). MODIS Land Surface Temperature Products Users ' Guide. Technical report, ICES, University of California, Santa Barbara, Santa Barbara.
- Wan, Z. (2008). New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. *Remote Sensing of Environment*, 112(1):59–74.
- Wan, Z. and Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on geoscience and remote sensing*, 34(4):892–905.
- Wan, Z. and Li, Z.-L. (2008). Radiance-based validation of the V5 MODIS land-surface temperature product. *International Journal of Remote Sensing*, 29(17-18):5373–5395.
- Wan, Z., Zhang, Y., Zhang, Q., and Li, Z.-L. (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment*, 83:163–180.
- Wan, Z., Zhang, Y., Zhang, Q., and Li, Z. L. (2004). Quality assessment and validation of the MODIS global land surface temperature. *International Journal of Remote Sensing*, 25(1):261–274.
- Wang, K. and Liang, S. (2009). Evaluation of ASTER and MODIS land surface temperature and emissivity products using long-term surface longwave radiation observations at SURFRAD sites. *Remote Sensing of Environment*, 113(7):1556–1565.
- Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., Zhou, X., and Haginoya, S. (2005). Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products. *Journal of Geophysical Research*, 110(D11):D11109.

- Wang, W., Liang, S., and Meyers, T. (2008). Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sensing of Environment*, 112(3):623–635.
- Yu, Y., Tarpley, D., Privette, J. L., Flynn, L. E., Xu, H., Chen, M., Vinnikov, K. Y., Sun, D., and Tian, Y. (2012). Validation of GOES-R Satellite Land Surface Temperature Algorithm Using SURFRAD Ground Measurements and Statistical Estimates of Error Properties. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3):704–713.

Appendices

A Global maps of AATSR for 15 July 2007 and 15 December 2007

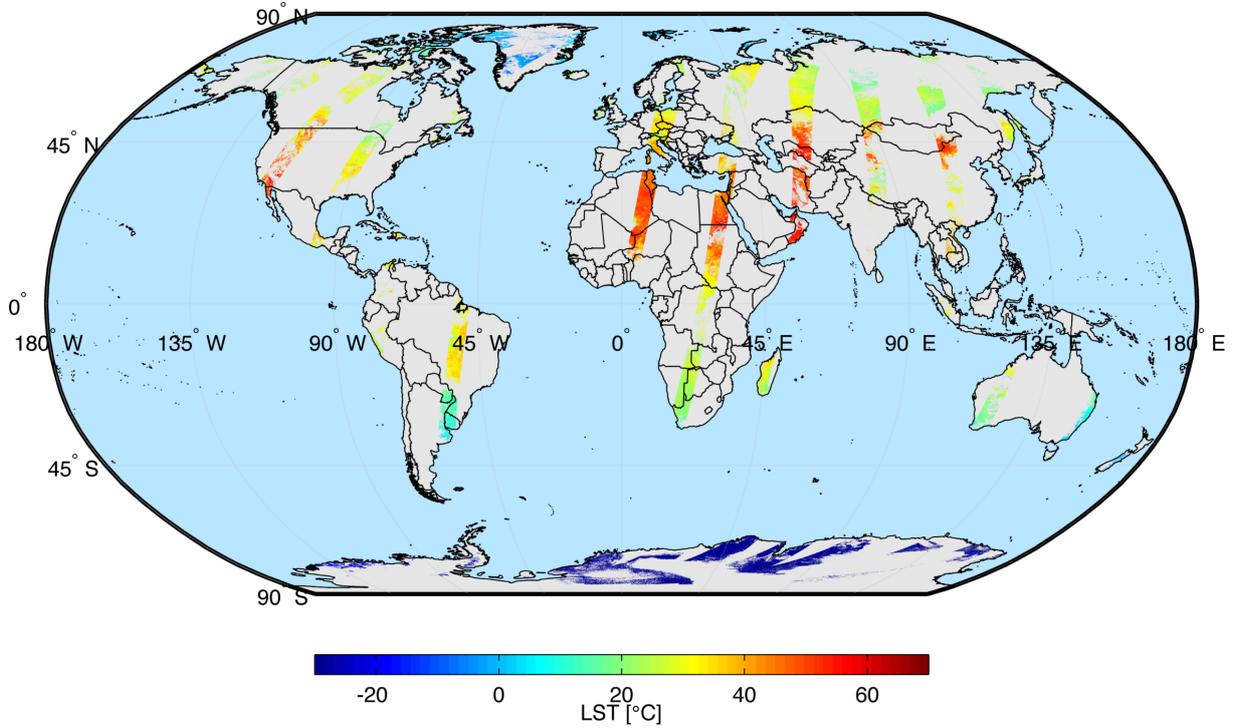


Figure 51 – Global mosaic of the AATSR product on 15 July 2007 for daytime data.

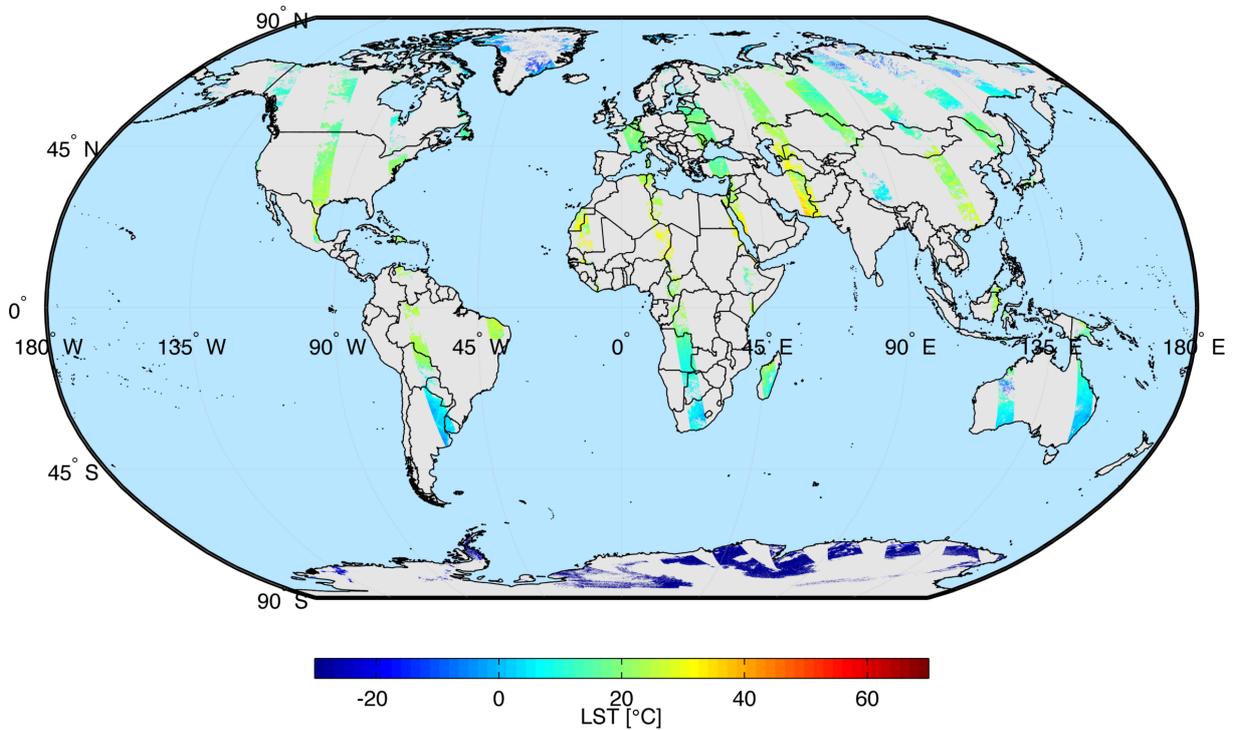


Figure 52 – Global mosaic of the AATSR product on 15 July 2007 for nighttime data.

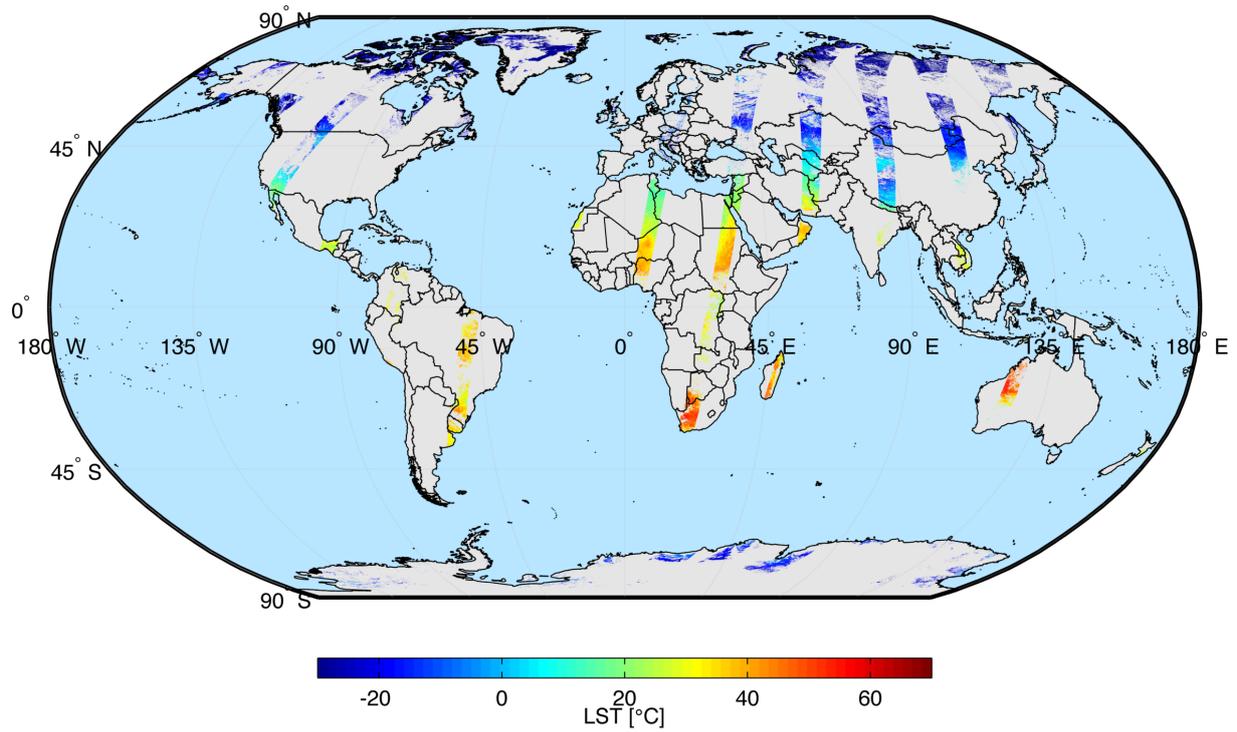


Figure 53 – Global mosaic of the AATSR product on 15 December 2007 for daytime data.

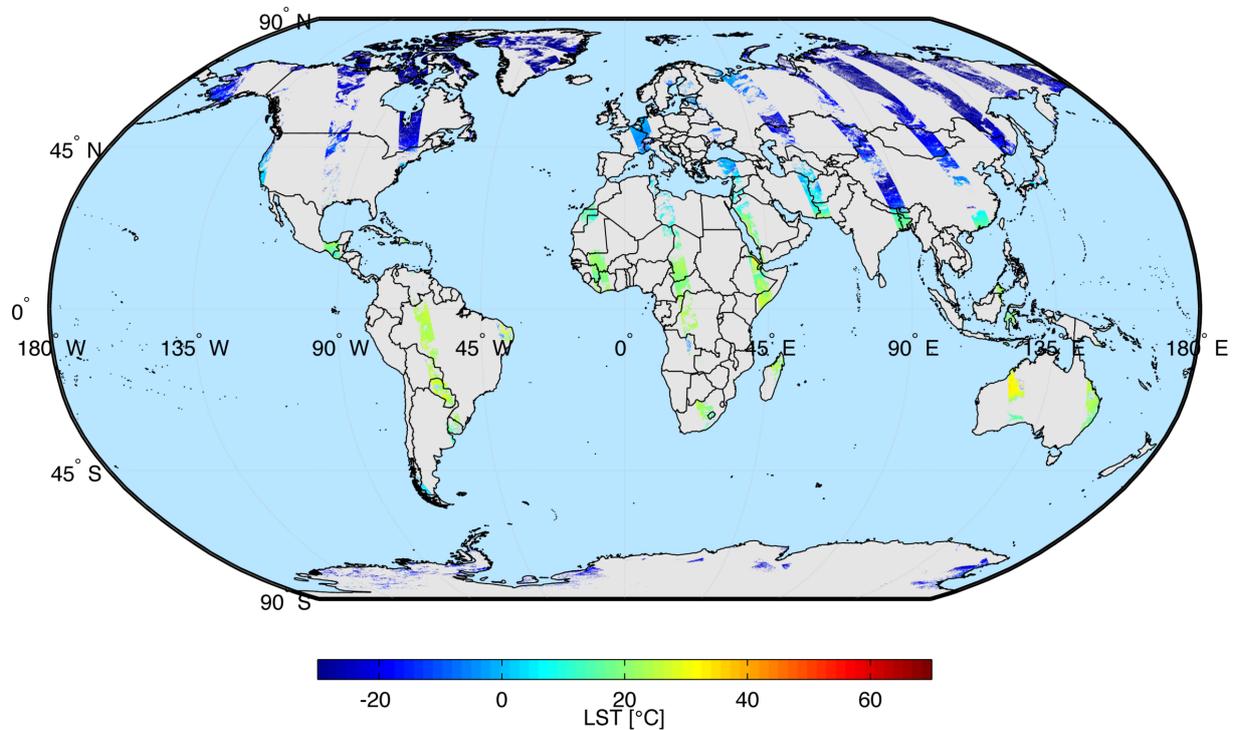


Figure 54 – Global mosaic of the AATSR product on 15 December 2007 for nighttime data.

B Global maps of MODIS for 15 July 2007 and 15 December 2007

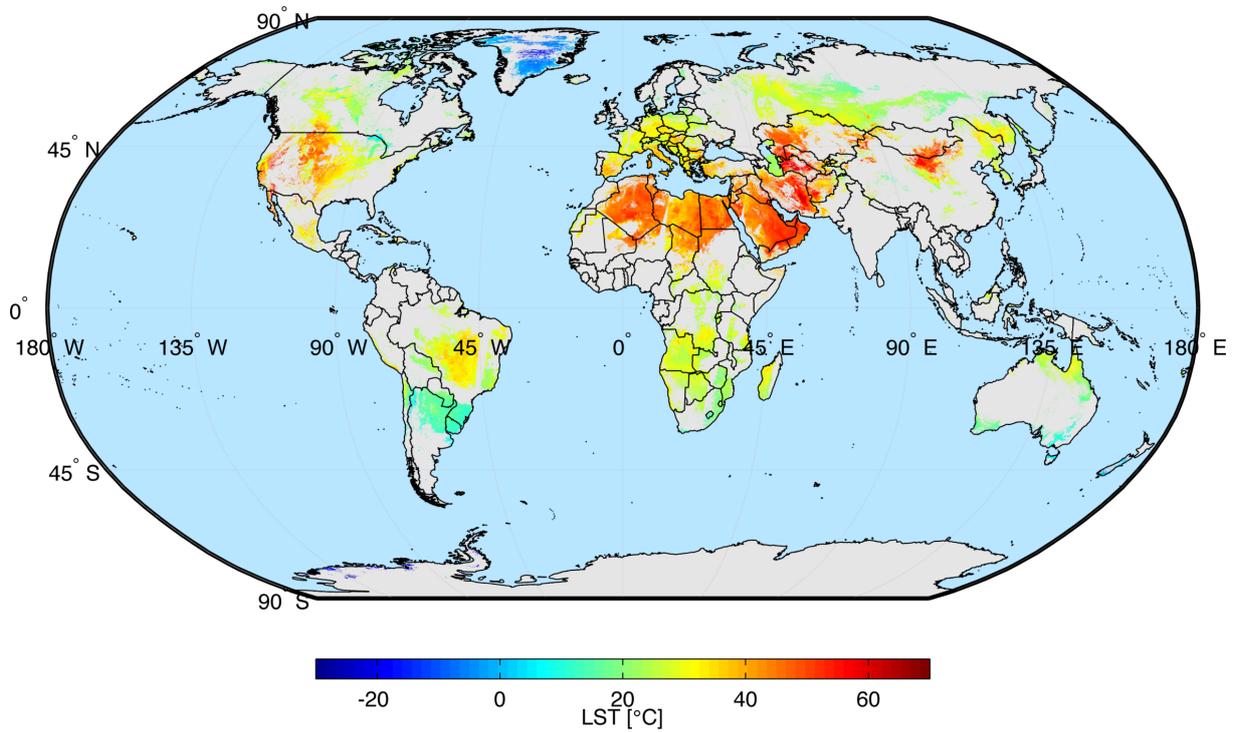


Figure 55 – Global mosaic of the MODIS MOD11A1 product on 15 July 2007 for daytime data.

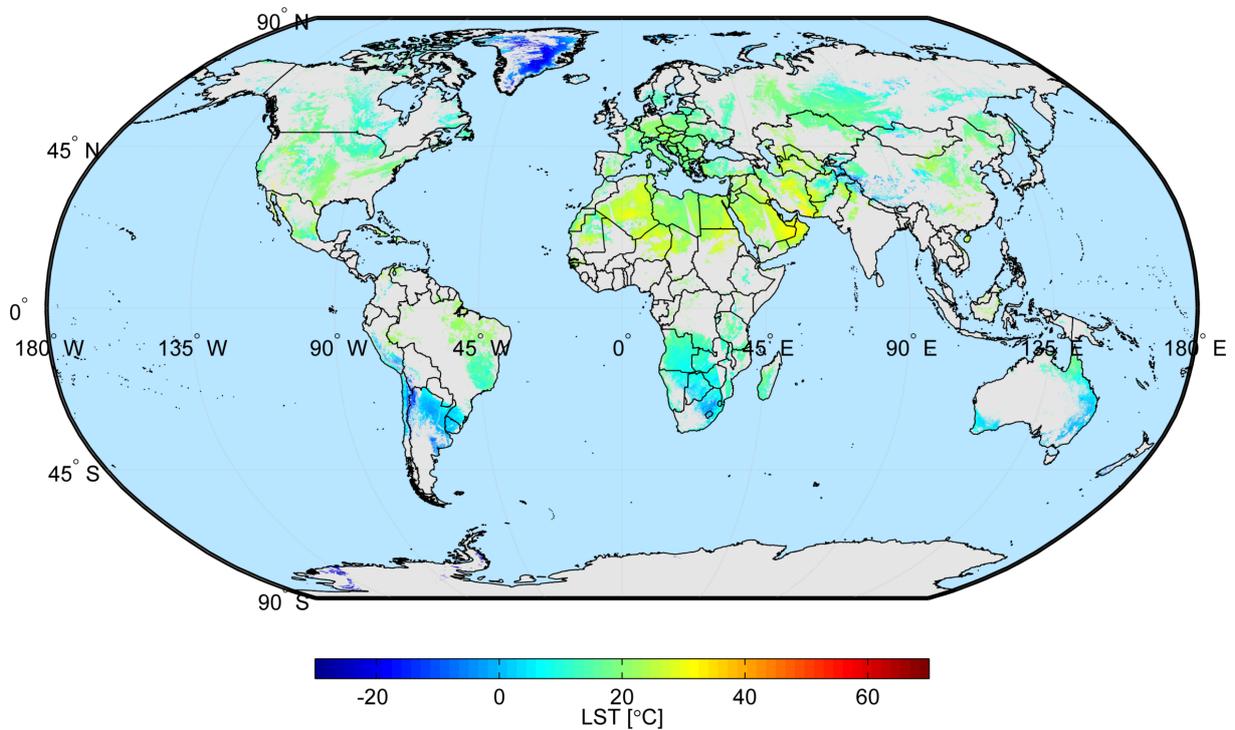


Figure 56 – Global mosaic of the MODIS MOD11A1 product on 15 July 2007 for nighttime data.

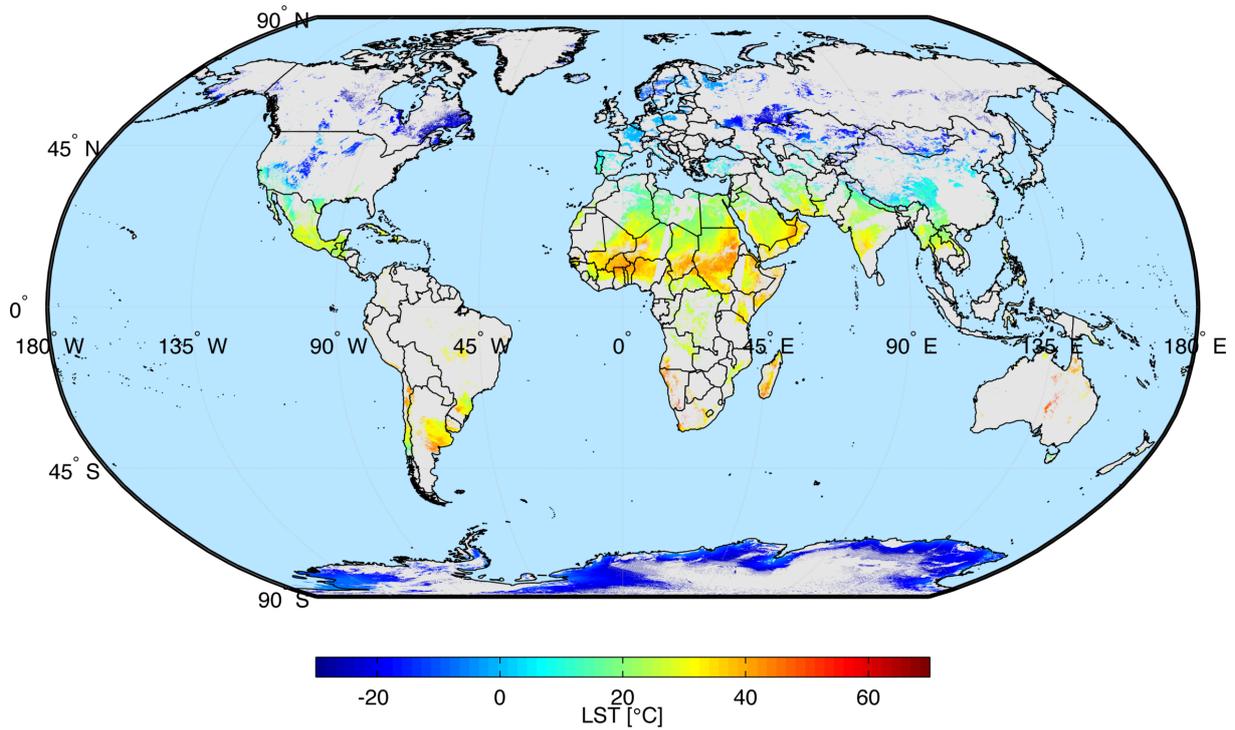


Figure 57 – Global mosaic of the MODIS MOD11A1 product on 15 December 2007 for daytime data.

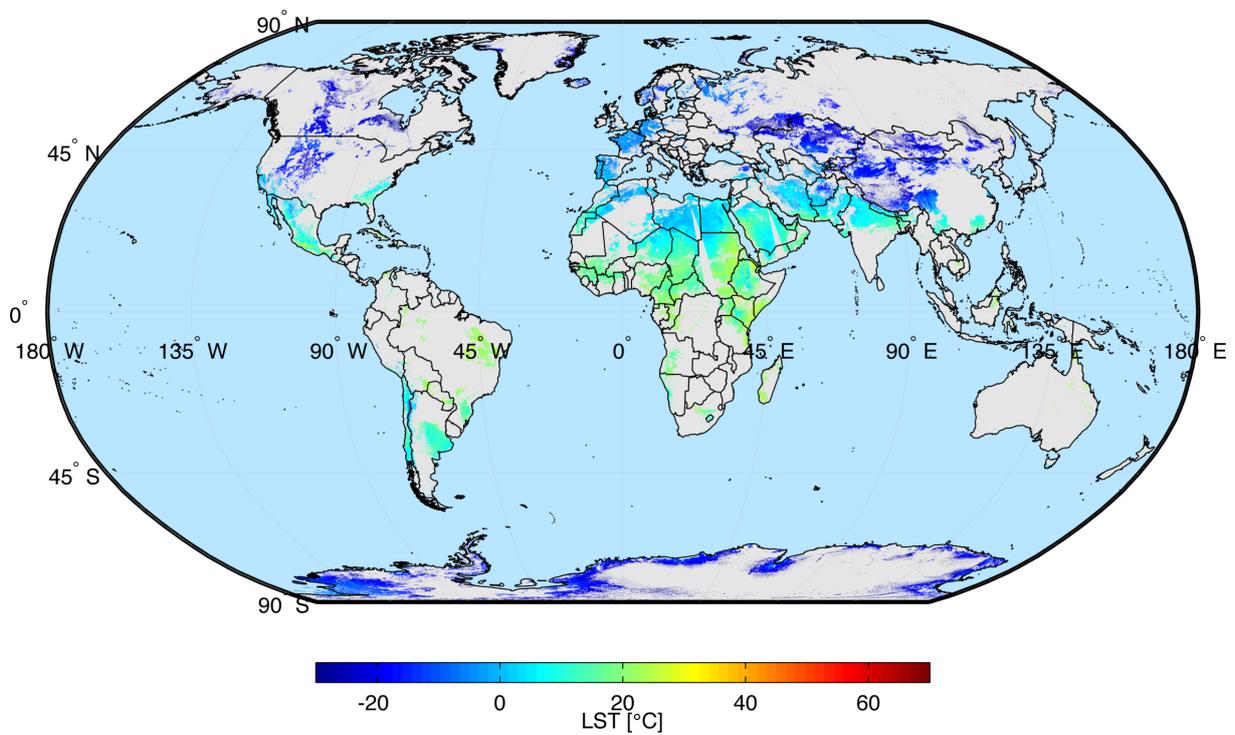


Figure 58 – Global mosaic of the MODIS MOD11A1 product on 15 December 2007 for nighttime data.

C Maps of MTSAT for Australia for 15 July 2007 and 15 December 2007

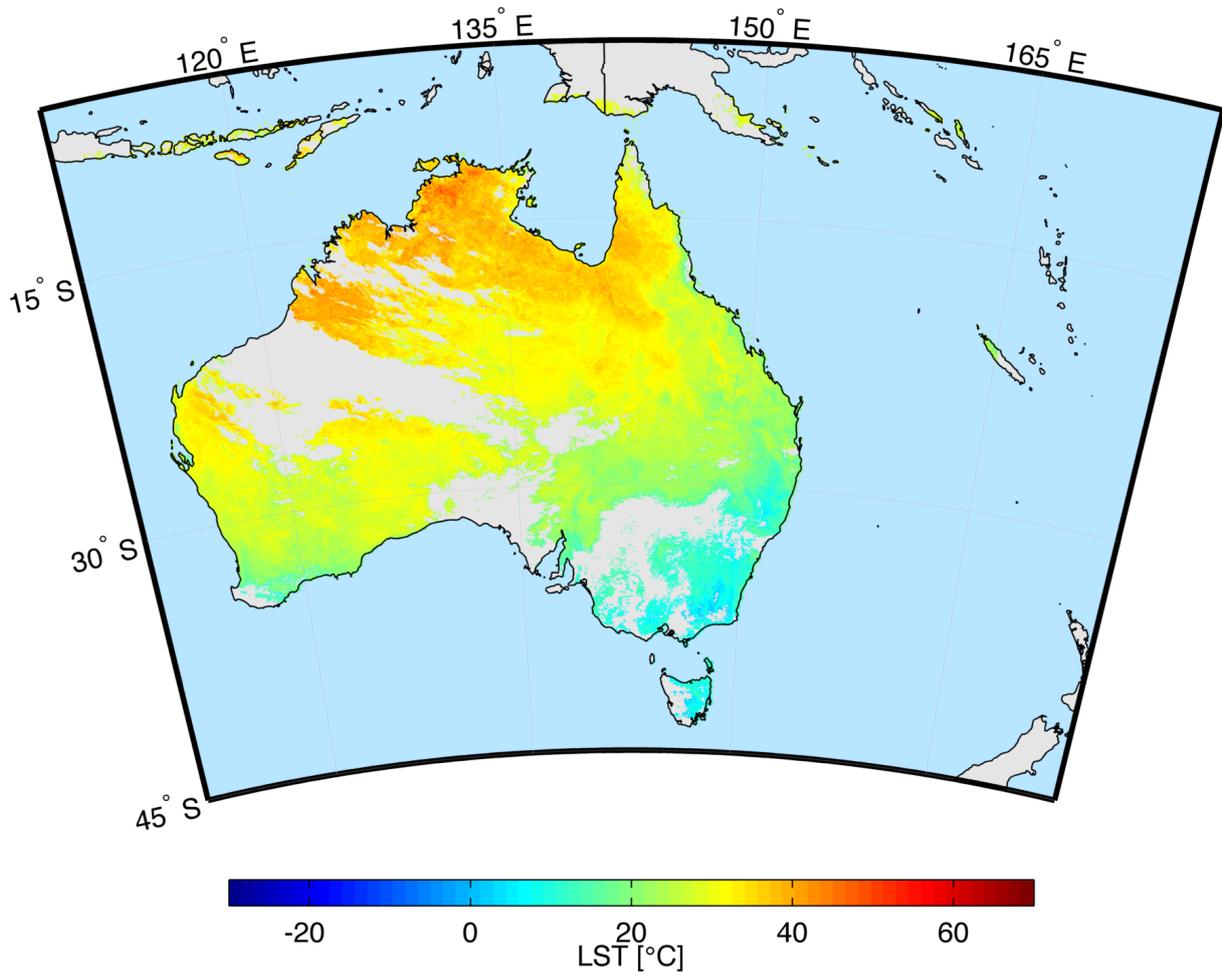


Figure 59 – Mosaic of the MTSAT LST product on 15 July 2007 for daytime data.

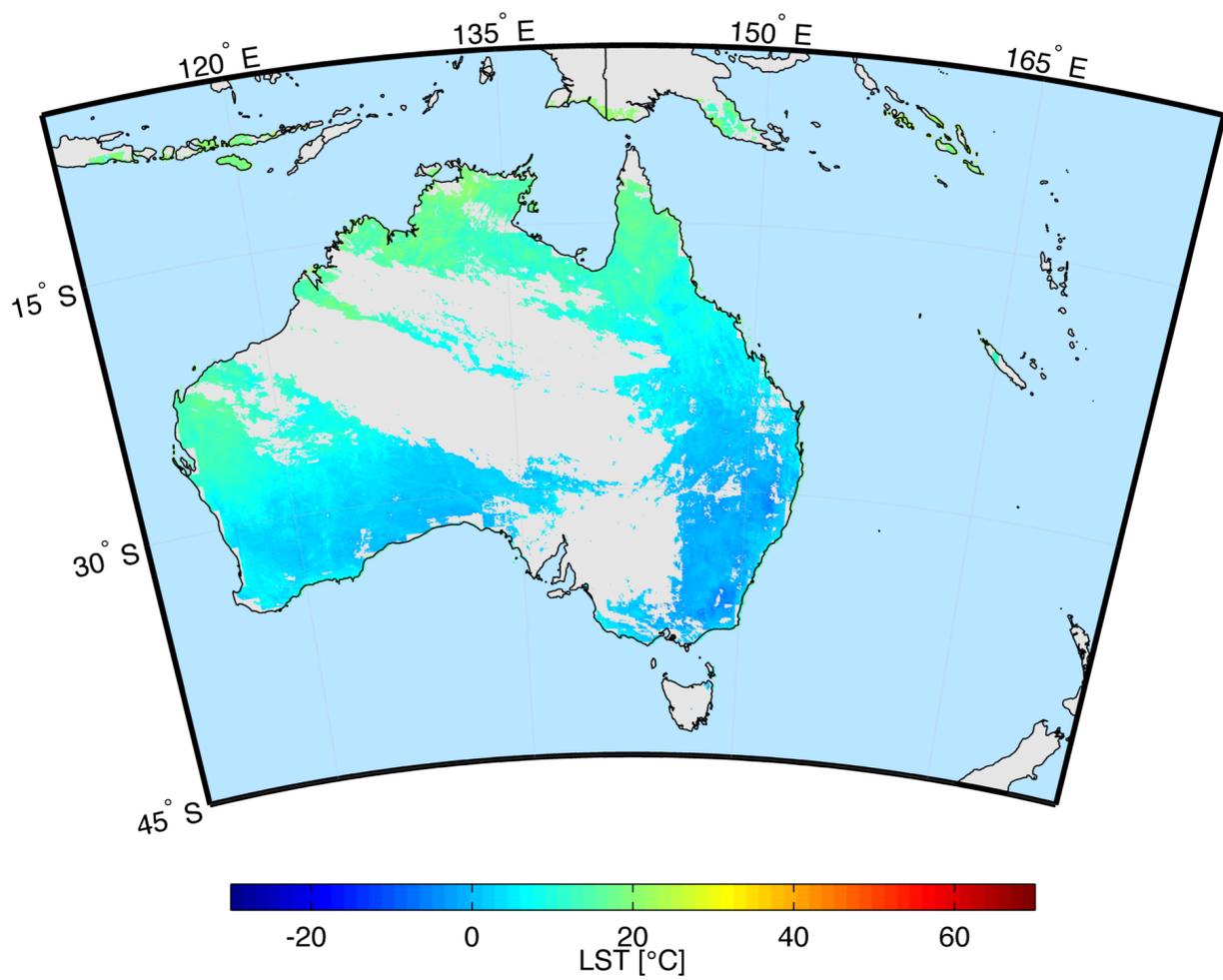


Figure 60 – Mosaic of the MTSAT LST product on 15 July 2007 for nighttime data.

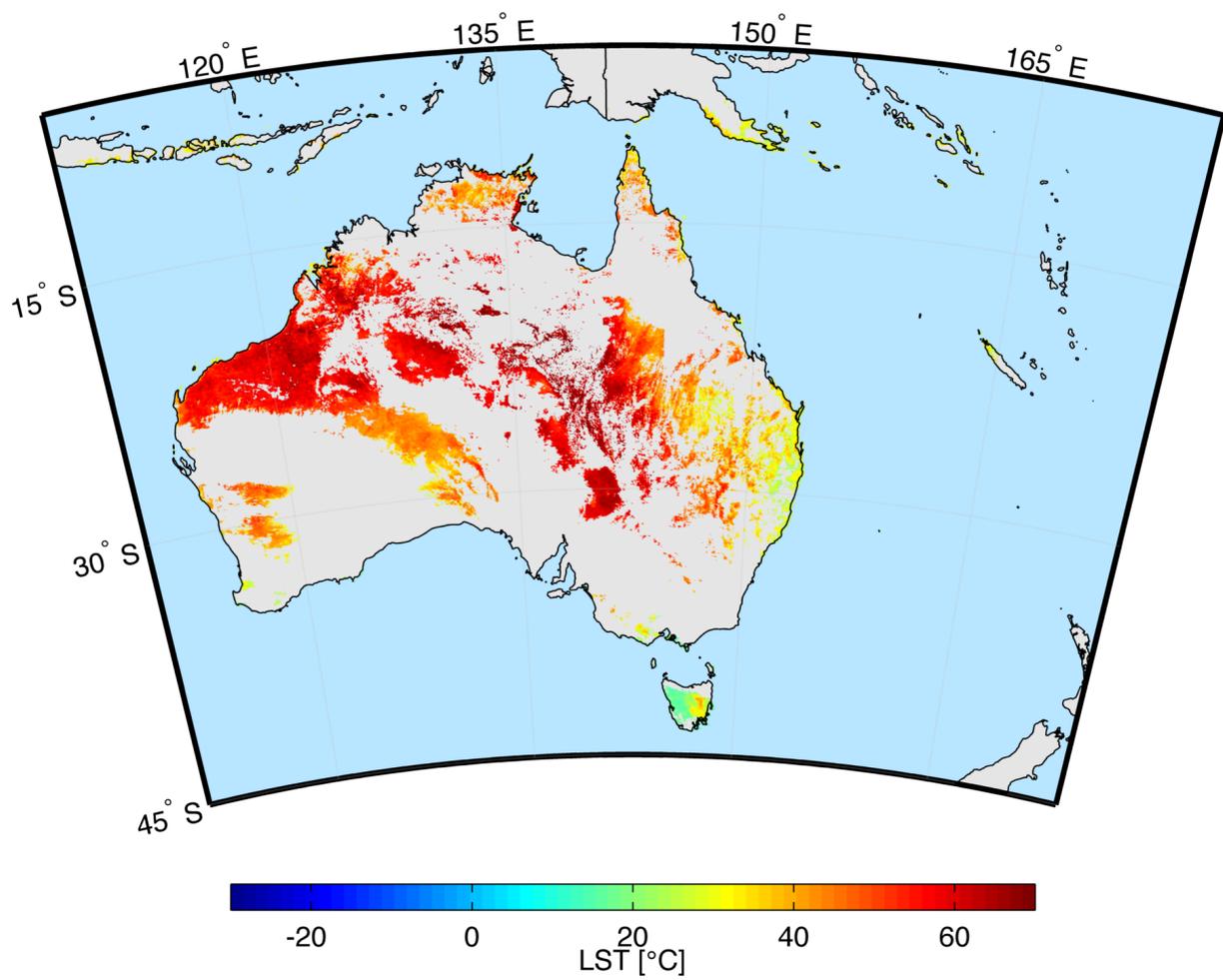


Figure 61 – Mosaic of the MTSAT LST product on 15 December 2007 for daytime data.

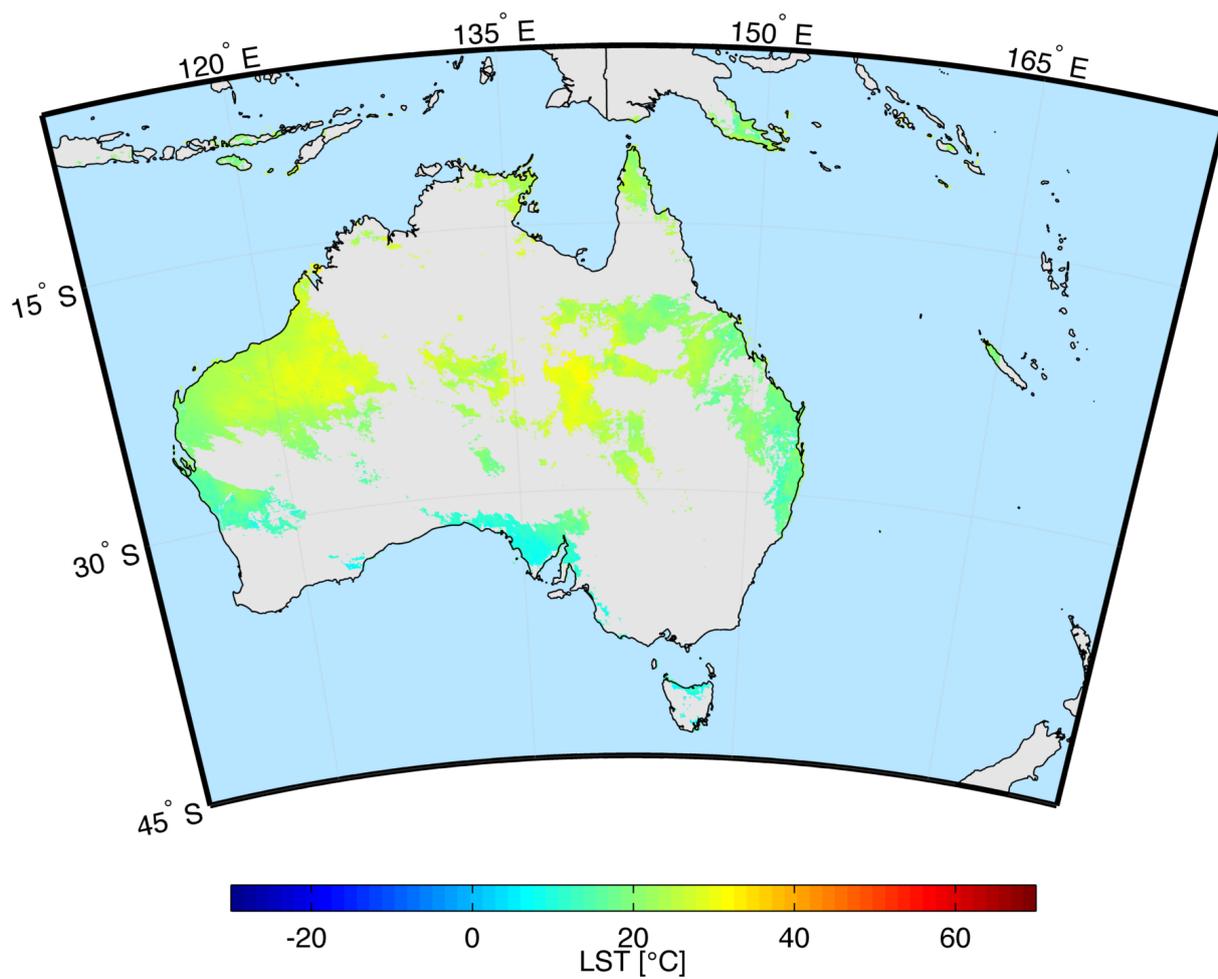


Figure 62 – Mosaic of the MTSAT LST product on 15 December 2007 for nighttime data.

D Maps of GOES-E for 15 July 2007

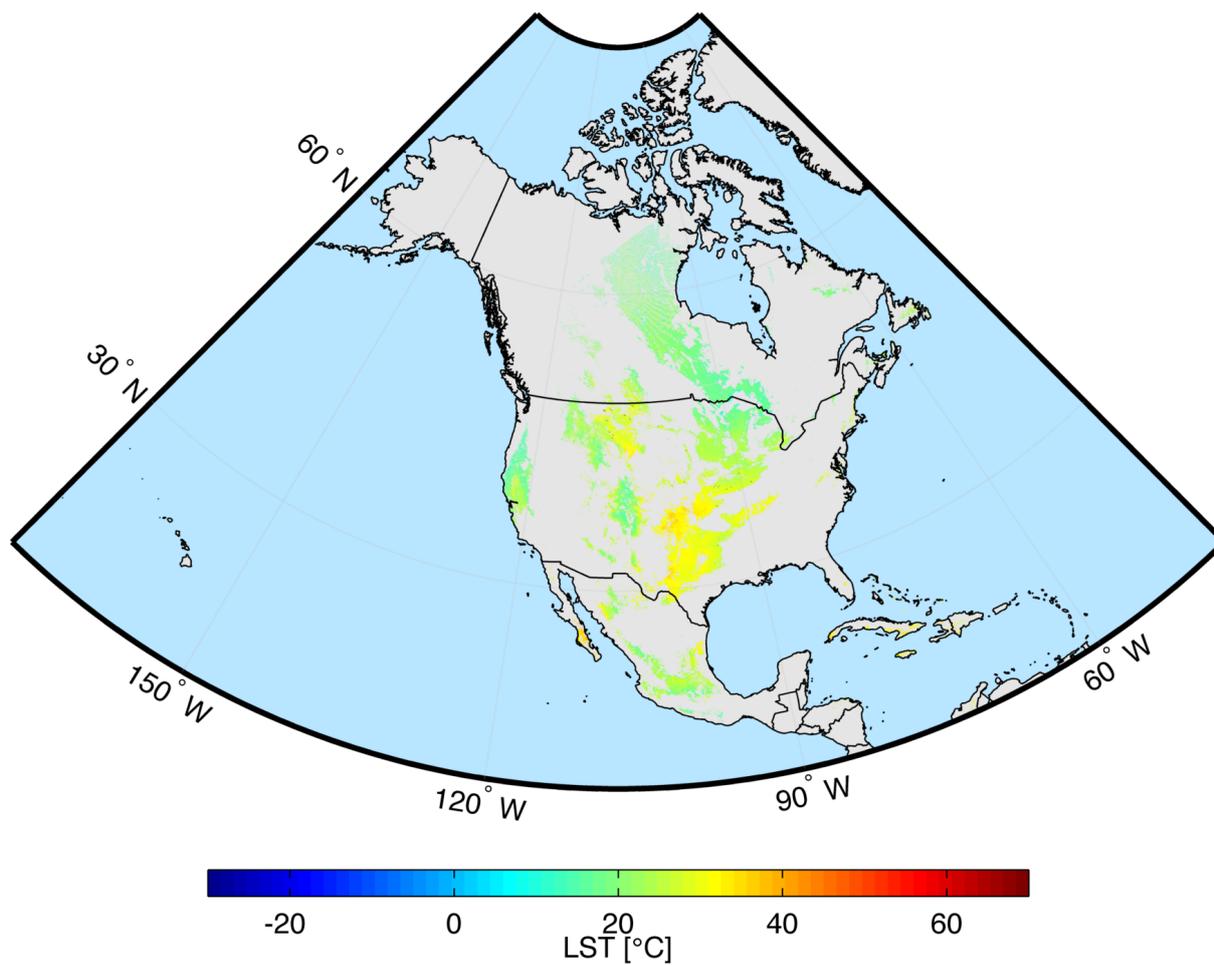


Figure 63 – Mosaic of the GOES-E LST product on 15 July 2007 for daytime data.

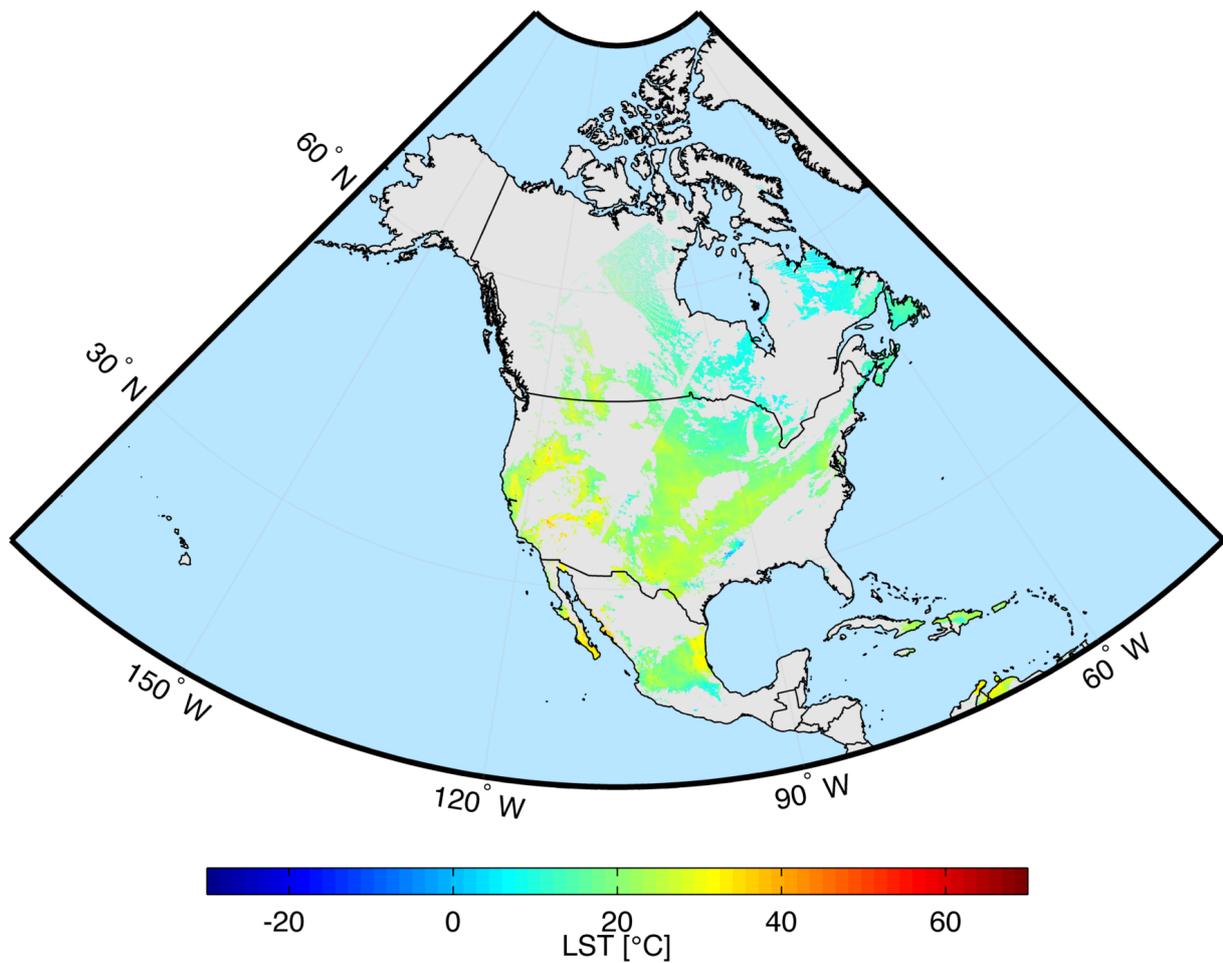


Figure 64 – Mosaic of the GOES-E LST product on 15 July 2007 for nighttime data.

NILU – Norwegian Institute for Air Research

NILU – Norwegian Institute for Air Research is an independent, nonprofit institution established in 1969. Through its research NILU increases the understanding of climate change, of the composition of the atmosphere, of air quality and of hazardous substances. Based on its research, NILU markets integrated services and products within analyzing, monitoring and consulting. NILU is concerned with increasing public awareness about climate change and environmental pollution.

NILU's values: Integrity - Competence - Benefit to society

NILU's vision: Research for a clean atmosphere

NILU – Norwegian Institute for Air Research
P.O. Box 100, NO-2027 KJELLER, Norway

E-mail: nilu@nilu.no

<http://www.nilu.no>

ISBN: 978-82-425-2876-6
ISSN: 2464-3327