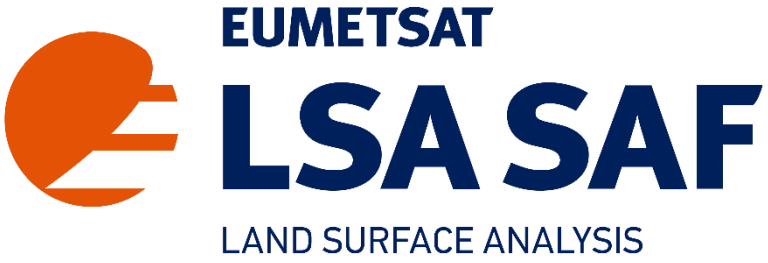
The EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA SAF)

VALIDATION REPORT
MLST-R (LSA-050)



Reference Number:	SAF/LAND/IPMA/VR_ MLST-R /v1.0
Issue/Revision Index:	Issue 2
Last Change:	11/12/2018

DOCUMENT SIGNATURE TABLE

	Name	Date	Signature
Prepared by :	Célia Gouveia, João P. A. Martins, I. F. Trigo, S. Coelho, F. Göttsche, F. Olesen		
Approved by :	LSA SAF Project Manager (IM)		

DOCUMENTATION CHANGE RECORD

Issue / Revision	Date	Description:
Version 1.0	24/07/2018	Version to be presented to DRR.
Issue 2	11/12/2018	- Introduction: clarifications about i) the climate applications of the MLST dataset; ii) issues related with the inclusion of Dahra Insitu station in the validation exercise; iii) rephrasing of some insitu and MODIS results from the validation exercise. - Validation methodology: add clarifications about bias and RMSE measure applications. - Adding data and information for Dahra insitu station, including row/column in Tables 3/5, new figures 5 and 9 and comments about bias and RMSE results. - Figures 6 and 10 replaced due to incorrect bias and RMSE values. - Figure 12 (with red squares representing the boxes used for decadal trend analysis) and clarifications about it in the text. - Figures 13, 14 and 15: improved readability. - Correct the number of new table 4 - Add comment about Tskin and MLST seasonality over IBE and SEN and about the lower daily amplitude over arid and semi-arid regions during the dry season observed for model surface temperatures (add of references) - Correction of names of regions for MODIS-LST and MLST comparison. - Rephrasing of some insitu and MODIS results from the validation exercise.

Executive Summary

Land Surface Temperature retrieved from SEVIRI/METEOSAT (LSA-001 product), is generated on an operational basis since February 2005 for the European region and since July 2005 for the whole METEOSAT disk. The main algorithm for LST estimation is based on a Generalized Split Window (GSW) that uses the difference between two adjacent window channels to correct the atmospheric absorption. Nevertheless, at the starting time of the operational generation, the GSW algorithm was not completely consolidated and several small modifications were implemented, namely an updated treatment of emissivity, upgrades in the ECMWF forecasts, among others. In the meantime, the generation of a long-term dataset was also pursued, aiming to include these datasets in climate variability and extremes monitoring studies. Consequently, LST was reprocessed for the period from 2004 to 2015 (LSA-050) with a temporal sampling of 15 minutes, using the version 7.14.0 for product LSA-001. The LST are freely available at http://lsa-saf.eumetsat.int.

This document presents validation results obtained for LSA-SAF Reprocessed LST retrieved from SEVIRI/METEOSAT (LSA-050, MLST-R) and, hereafter referred as MLST-R. The purpose of the validation report is to characterize the MLST-R dataset in terms of accuracy, precision and stability. The first goal of this validation process is focused on the determination of accuracy and precision of MLST-R in comparison in-situ LST obtained from ground measurements in validation sites. The second step of this validation exercise is to show the temporal stability of the MLST-R dataset. To achieve this goal, we performed an assessment for the period from 2004 to 2015, of the trend in bias, obtained over selected sites for the difference between MLST-R time series and i) ECMWF ERA-Interim skin temperature and ii) MODIS LST (Collection 6).

For validation purposes related to satellite-derived LST products over a wide range of surface and climatic conditions, Karlsruhe Institute of Technology (KIT) set up four permanent LST validation stations in areas characterized by naturally homogeneous land use and land cover in different climate zones (Evora, Portugal; Dahra, Senegal; Gobabeb and RMZ, both in Namibia).

In-situ validation exercises performed over Evora, Gobabeb and RMZ show the fulfilment of the optimal accuracy of 1 K is achieved for the 7 considered years in the 2 stations (Gobabeb and EMZ) for instantaneous, daily and monthly match-ups, except 2014 in Gobabeb. In terms of RMSE, the values are below 2 K for the all cases in both stations, except for Gobabeb in 2014 (instantaneous) and 2011 (daily). Nevertheless, RMSE never exceed 2.2 K. Evora fulfils the target accuracy of 2.0 K for all years and cases and in terms of RMSE the values are close to 2.0 K. However, if only night-time instantaneous match-ups are considered, then the optimal requirement is also achieved for Evora; spatial heterogeneity tends to be the largest during daytime. Dahra fulfills the 2.0 K accuracy requirement in 2014 and 2015 and the 4.0 K requirement in all years with exception of 2009.

A set of 10º longitude x 10º latitude areas (Iberian Peninsula, Central Europe, Algeria, Savana, Senegal and Gobabeb) are chosen to include in situ stations used to validate LST_SEVIRI and to embrace a wide range of biomes and different atmospheric conditions within the MSG disk. The decadal stability is assessed by means of computing the bias between MLST-R and the ECMWF ERA-Interim skin temperature for a window of 0.75º x 0.75º centred in the above mentioned areas and corresponding linear trend. The results indicate the latter is within 0.2 K/decade at two (Senegal and Iberia) out of six investigated locations and within 0.4 K/decade for in Central Europe and Algeria

boxes. On the other hand, Gobabeb and Savana exhibit higher decadal trend values with opposite signs (positive for Savana and negative for Gobabeb), although not significant.

The decadal stability of the MLST-R was also evaluated against the Level-2 v006 MOD11_L2 and MYD11_L2 from Terra and Aqua, over a 1° x 1° latitude and longitude window centred at the six defined areas. A trend of 0.4 K/decade is observed in Savana and Iberia areas, respectively for day and night-time, while better performance of 0.2 K/decade is reached at four out of six investigated locations for both day- and night-time match-ups. Furthermore, the decadal trends obtained were not significant (significance level above 5%) over Algeria, Iberia, Savana and Senegal for day-time match-ups and over Algeria, Central Europe and Iberia for night-time match-ups, being negative for all the regions, with exception of Gobabeb for day-time and Algeria for night-time match-ups.

TABLE OF CONTENTS

List of Tables

List of Figures

1 Introduction

LSA-SAF retrieves Land Surface Temperature (LST) from directional surface emitted Thermal Infrared radiances (TIR) that are derived from cloud free measurements from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) on Meteosat Second Generation (MSG) on an operational basis (LSA-001). LST is assumed to be the radiative skin temperature of the land surface that emits IR radiance (Norman and Becker, 1995), as measured in the direction of the sensor. Directional radiometric temperature is assumed to be the best estimate of thermodynamic temperature that can be obtained from a radiance quantity. The LSA-SAF LST estimates from SEVIRI follow the 'Generalised Split-Windows' (GSW) approach, proposed by Wan and Dozier (1996) to estimate LST from MODIS. As such, LST is estimated within the LSA-SAF as a linear function of clear-sky top of the atmosphere (TOA) brightness temperatures from channels 10.8 and 12.0 μm on SEVIRI, obtained by the split-window formulation, assuming surface emissivity known for both bands (Trigo et al., 2008a). The estimation of the GSW parameters relies on linear regressions of synthetic brightness temperatures, obtained from radiative transfer simulations (using MODTRAN4, Berk et al., 2004) over a wide range of surface and atmospheric conditions.

In 2017 LST for the period between 2004 to 2015 (LSA-050) was reprocessed by LSA-SAF with a temporal sampling of 15 minutes, using the version 7.14.0 for product LSA-001. The LST data generated by the LSA SAF are freely available from http://lsa-saf.eumetsat.int.

This document presents validation results obtained for LSA-SAF Reprocessed LST retrieved from SEVIRI/METEOSAT (LSA-050, MLST-R) and, hereafter referred as MLST-R.

2 Validation Methodology

The purpose of the validation report is to characterize the MLST-R dataset in terms of accuracy (as obtained using bias), precision (as obtained using RMSE) and stability (as obtained computing decadal trend of bias). Furthermore, the products are confronted with the product requirements stated in Product Requirements Document version 3.0, SAF/LAND/PRD/3.0 and their compliance is reported. These requirements are summarized in Table 1. Usually only bias is used as a measure of accuracy (Table 1), however in this validation exercise we will also assess the RMSE, which accounts for random and systematic errors.

Table 1 Product Requirements for LSA SAF LST Reprocessed, LSA-050, in terms of area coverage, resolution and accuracy (Product Requirements Document version 3.0, SAF/LAND/PRD/3.0.

LST Product	Coverage	Resolution		Accuracy		
		Temporal	Spatial	Threshold	Target	Optimal
MLST-R (LSA-050):	MSG disk	15 min	MSG pixel resolution	4K	<2 K	1K

The first goal of this validation process is focused on the determination of accuracy and precision of our climate record dataset, i.e., to show a measure of proximity of MLST-R in comparison with the 'truth' data. For this assessment, we will use in-situ LST obtained from ground

measurements in validation sites which are the most conclusive and independent datasets for validating remotely sensed LST (Schneider et al., 2012).

LST is a variable that presents high variability in space and time. Therefore, the selection of the validation sites aims to cover a wide range of atmospheric conditions on the Meteosat disk. However, only the LSA SAF sites located in highly homogeneous areas were used as they provide meaningful accuracy and precision estimates for satellite-based LST retrievals (Göttsche et al. 2016). In order to achieve this goal, the bias and root mean square error (RMSE) between MLST-R and in-situ LST were computed. Those accuracy measures were computed excluding outliers obtained by the application of a filter of 3σ (Göttsche et al. 2016, Schneider et al., 2012).

The validation of LST retrieved by satellite using in-situ LST values must be considered carefully due to the low number of high-quality field datasets and the lack of global representativeness. The high thermal heterogeneity introduced by the different characteristics of the surface brings several types of uncertainties, namely those related with the upscaling of the in-situ measurements to satellite pixel size (Göttsche et al., 2013; Ermida et al., 2014; Jimenez-Munoz et al., 2014). We provide validation statistics for instantaneous, daily and monthly data, where the latter are computed using the original 15-min MLST-R product. To account for time differences between the Meteosat acquisitions and the reference data sets, collocations are first rendered at instantaneous level, and then aggregated.

In the case of the in-situ measurements for Evora, a correction for geometric effects were performed, taking in account the results of Ermida et al. (2014), where the upscaling of ground measurements takes into account illumination and viewing angles.

The second step of this validation exercise consists in the estimation of the temporal stability of MLST-R dataset for the period between 2004 and 2015. Decadal stability may be viewed at as the change of LST accuracy (as obtained using bias) in time (i.e. per decade). This exercise will be performed by comparing the stability of MLST-R against two different datasets: i) ECMWF ERA-Interim skin temperature; and ii) MODIS-LST (Collection 6).

This validation report is applicable to MLST-R product generated by the LSA SAF operational chain (MLST-R).

3 Accuracy and Precision

3.1 In-situ LST validation stations

For validation purposes related to satellite-derived LST products over a wide range of surface and climatic conditions, Karlsruhe Institute of Technology (KIT) set up four permanent LST validation stations in areas characterised by naturally homogeneous land use and land cover in different climate zones. Although being part of LSA-SAF's validation effort and chosen to validate LST derived from MSG/SEVIRI, the validation stations are equally well suited to validate LST products from other sensors. Table 2 presents the characteristics of the geo-climatic stations used in this validation exercise and Figure 1 shows their locations within the field of view (FOV) of the METEOSAT satellites.

In-situ measurements at Evora started in 2005, but the set-up at the current location and orientation of radiometers is in operation since March 2009. Bare ground is usually not observed at Evora validation station (Figure 2), even in August the surface is covered by dry and relatively high grass, which increases effective emissivity via the cavity effect (internal reflection within the plant canopy leading to the increase of land surface emissivity in the thermal infrared with the amount of

vegetation) (French et al., 2000; Olioso et al., 2007). Furthermore, a 9-minute delay between actual satellite acquisition time for Evora (SEVIRI scans the Earth from South to North) and nominal product time has been accounted for and in-situ LST and MLST-R were matched to better than 5 minutes.

Table 2 Main geo-climatic characteristics of KIT's validation stations used in this validation report.

Stations Name	Country	Latitude	Longitude	Start date	Altitude (a.s.l.)	Climate zones (Köppen, 1936)	Land cover type
Evora	Portugal	38.540ºN	8.003ºW	2005 (2009)	230	Mediterranean climate (CSh)	Cork-oak trees/ grass
Gobabeb	Namibia	23.551°S;	15.051°E	2007	450	Warm desert climate (BWh)	Gravel plain
RMZ farm/ farm Heimat	Namibia	22.933°S;	17.992°E	2009	1380	Warm desert climate (BWh)	Kalahari bush
Dahra	Senegal	15.402°S	15.433°E	2009	90	Hot semi-arid climate (BSh)	Grassland

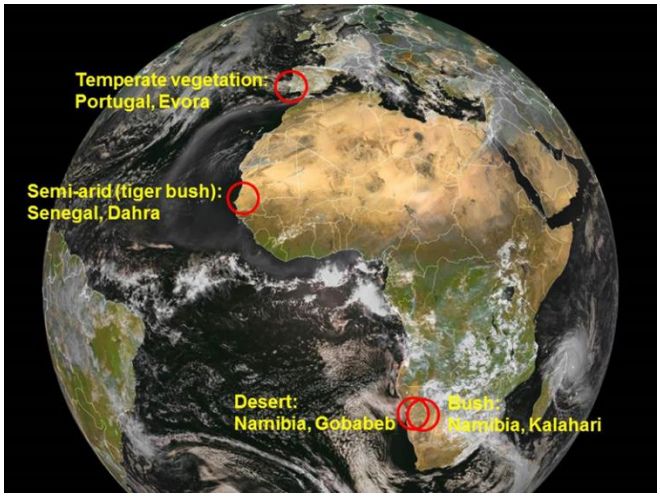


Figure 1 Locations of KIT's validation stations on MSG/SEVIRI earth disk.

Figure 2 Landscape of the location of Evora validation station.

Gobabeb LST validation station lies about 2 km north-east of Gobabeb Training & Research Centre (www.gobabebtrc.org) in the Namib Desert, Namibia (Figure 3). A sharp transition between the Namib sand sea with its up to 300 m high dunes and the gravel plains is observed and maintained by irregular flows of the transient Kuiseb River (few days every year) which wash the advancing sand into the South Atlantic Ocean. The site is suitable for long-term validation studies of satellite products due to the high spatial and temporal stability typical of hyper-arid desert climate (Peel et al, 2007, Hulley et al., 2009). The high average annual temperature (21.1°C, Lancaster et al., 1984) and very low and highly variable average annual precipitation (less than 100 mm) highlights the role of relatively frequent fog events in regional water balance (Peel et al., 2007, Eckardt et al., 2013).

Figure 3 Landscape of the location of Gobabeb validation station.

Farm Heimat lies about 100 km south-east from Windhuk on a plateau in the Kalahari semi-desert (Figure 4). The Kalahari is characterised by hot and arid climate presenting a short rainy season (very little rain from September to November) and a strong rainy season (January and March) with possible flooding. During the dry season the Kalahari bush is dry and the grass desiccates quickly. Farm Heimat produces livestock (cattle & sheep) carefully managed and moved systematically between fenced off 'camps' to avoid overgrazing. The station is in a typical Kalahari land scape and a wide area around the mast is mainly covered by patchy, desiccated grass dotted

with bushes and isolated camel thorn trees. The winter temperatures (June - August) frequently drop well below freezing point due to the high altitude and the remaining water vapour column between the surface and satellite is quite small (50% of the atmosphere's water vapour is limited in the lowest 2 km).

A fourth LST validation station is located about 7 km north-east of the town of Dahra, Senegal (Figure 5). The station is located in a practically unpopulated area, covered by seasonal grass (95%) and sparse trees. In some cases, the distribution of the bushes and trees follows ancient dunes, which causes stripes of high vegetation - hence the name "tiger bush". This validation site exhibits a strong annual vegetation cycle and also a strong intra annual variability of atmospheric water vapour with a very warm (about 40° C) and humid rainy season (up to 90 % relative humidity) that highlights the role of an accurate atmospheric correction of satellite TIR. Cloud detection becomes also difficult during occasional outbreaks of Sahara dust.

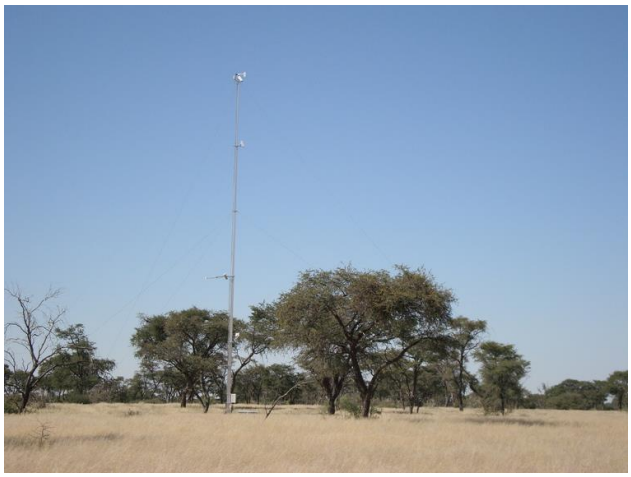


Figure 4 Landscape of the location of Farm Heimat validation station.

Figure 5 Landscape of the location of Dahra validation station.

The validation of LST remote sensing products using in-situ observations is still a stimulating problem due to the high spatial and temporal variability of LST. Attention should be devoted to daytime results over complex and structured land surfaces (e.g. the cork-oak tree forest at Evora station), which exhibit large thermal gradients between different land surface covers and where the dependence of their fractions on viewing and illumination geometry may have to be considered (Ermida et al., 2014; Guillevic et al., 2013). On the other hand, the validation of satellite LST and emissivity products and associated limitation of uncertainties of ground-based LST observations require accurate estimations of land surface emissivity (LSE) and continuous measurement of down welling radiance, namely in sites with larger fractions of bare soil. In-situ measurements performed at Gobabeb revealed that LSE estimations over arid regions can be incorrect by more than 3% (Göttsche and Hulley, 2012), leading to LST errors typically between 1°C and 2°C (Schädlich et al., 2001). To minimize such errors, in-situ LSEs of the dominant surface cover types at Gobabeb were obtained using the so-called 'emissivity box method' (Rubio et al., 1997) and from emissivity spectra of soil samples (Göttsche and Hulley, 2012). Further details about in-situ measurements and LST determination can be found in Göttsche and Hulley (2012).

3.2 In-situ validation results

Table 3 summarizes the number of instantaneous, daily and monthly valid points obtained from the match-ups between satellite and in-situ LST for Evora, Gobabeb, RMZ Heimat and Dahra from 2009 to 2015. For each year, the number of instantaneous valid points are mainly larger than 10.000, with exception of Evora and RMZ in 2014 (respectively 9393 and 5627 match-ups) and Dahra in 2009, 2010 and 2012 (respectively 5784, 7533 and 4679 match-ups). Concerning the daily and monthly valid points the picture is similar, presenting values higher than 200 days and 8 months, respectively. The exceptions are again Evora in 2013 (respectively 169 and 6), 2014 (respectively 185 and 7) and 2015 (respectively 85 and 4), RMZ in 2014 (respectively 105 and 6) and Dahra in 2009, 2010 and 2012 (respectively (117 and 4; 148 and 6; 1001 and 4).

Figure 6 to Figure 9 show plots of MLST-R derived from MSG/SEVIRI against LST retrieved from in situ data at annual basis for the period from 2009 to 2015 for Evora, Gobabed, RMZ Heimat Farm and Dahra respectively. During 2009 there are 13834, 22942, 16586 and 5784 match-ups between satellite and in-situ LST, which yielded biases (RMSE) of 1.4 K (2.4 K), -0.6 K (1.5 K), 0.2 K (1.9 K) and -4.2 K (5.4 K) respectively for Evora, Gobabeb, RMZ and Dahra. These overall statistics suggest the LSA SAF LSA-050 product meets its target accuracy (as obtained using bias) of 2 K for all in-situ stations with exception of Dahra.

Table 3 Number of instantaneous, daily and monthly valid points obtained after the match-ups between satellite and in-situ LST (N) for KIT's validation stations on MSG/SEVIRI earth disk from 2009 to 2015.

Year	Sampling	Evora	Gobabeb	RMZ Heimat	Dahara
2009	instantaneous	13834	22942	16586	5784
	daily	239	327	232	117
	monthly	10	12	9	5
2010	instantaneous	17179	26677	23620	7533
	daily	303	361	325	148
	monthly	12	12	10	6
2011	instantaneous	15901	20158	19923	13051
	daily	255	273	291	235
	monthly	10	11	11	11
2012	instantaneous	17981	19923	25335	4679
	daily	270	352	208	101
	monthly	10	12	9	4
2013	instantaneous	11186	23055	21788	18140
	daily	169	301	295	298
	monthly	6	11	11	11
2014	instantaneous	9393	23706	5627	21098
	daily	185	319	105	272
	monthly	7	12	6	10
2015	instantaneous	4246	27091	19635	17785
	daily	85	358	272	336
	monthly	4	12	10	12

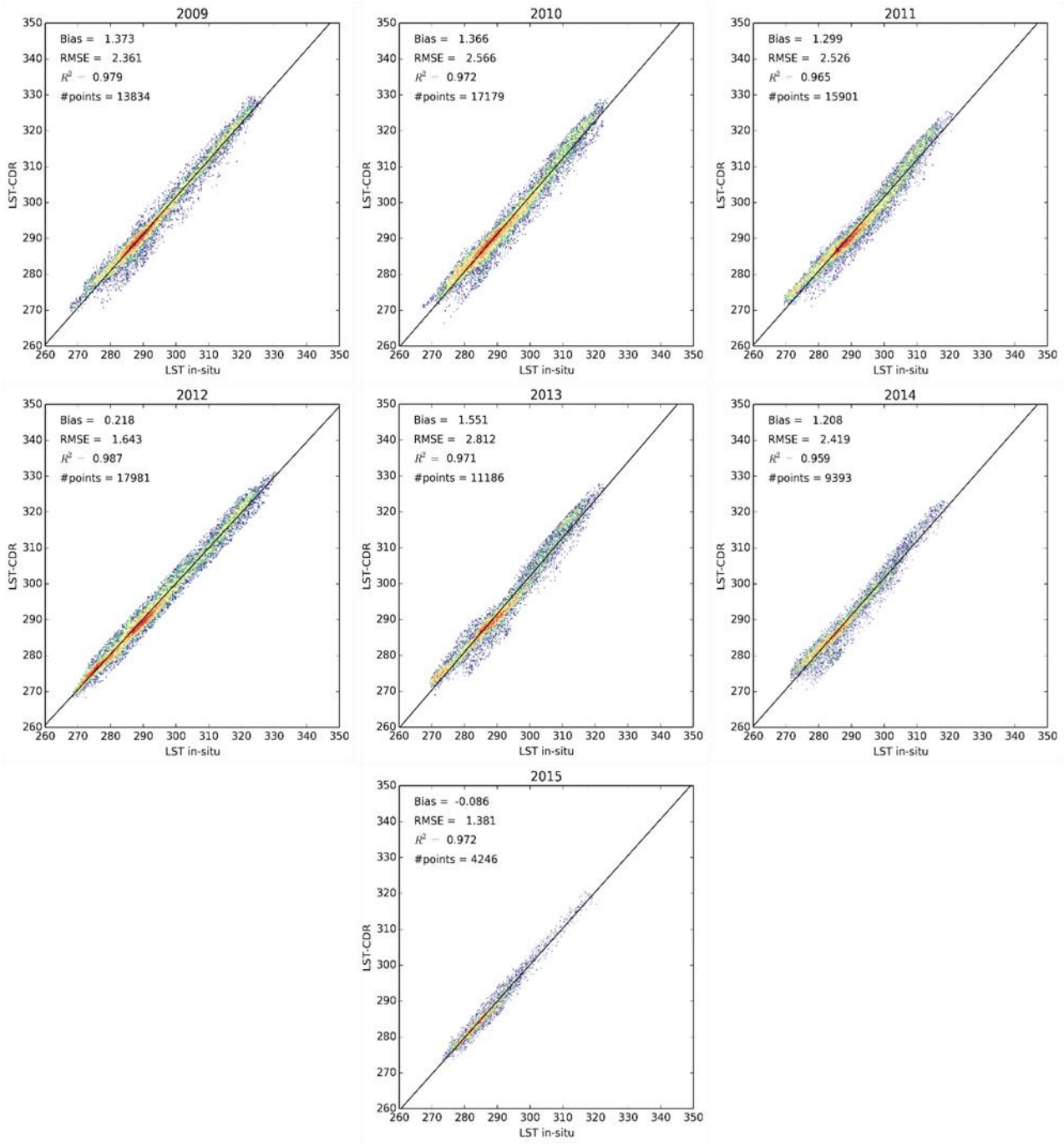


Figure 6 MLST-R (K) from MSG against in-situ LST (K) for Evora, at annual basis from 2009 to 2015.

Validation years 2010 and 2011 in Evora present similar results in terms of bias and RMSE (1.4 K and 2.5, respectively in both years) and 17179 (15901) match-ups between satellite and in-situ information. Better results were obtained for 2012 with bias and RMSE of 0.2 K and 1.6 K, respectively. During 2013, 2014 and 2015 larger bias and RMSE values are obtained (bias: 1.6 K, 1.2 K and -0.1 K; RMSE: 2.8 K, 2.4 K and 1.4 K), although still meeting the target accuracy (bias values). It is worth noting the lower number of match-ups between satellite and in-situ information in 2013 and 2014 (11186 and 9393), corresponding to a higher cloud cover. The statistics fulfil the optimal

requirement (1.0 K) for 2012 and 2015 and the target threshold of 2.0 K for all years. However, better accuracy results are expected when only night -time statistics are considered, as surface heterogeneities are greatly smoothed during that period (Trigo et al, 2008; Guillevic et al., 2013; Ermida et al, 2014). As such, the bias for night-time match-ups fulfils the optimal criteria for all years except 2009 (bias values ranging between -0.2K in 2012 and 1.1K in 2009; and RMSE from 1.0 in 2015, to 2.1K in 2014).

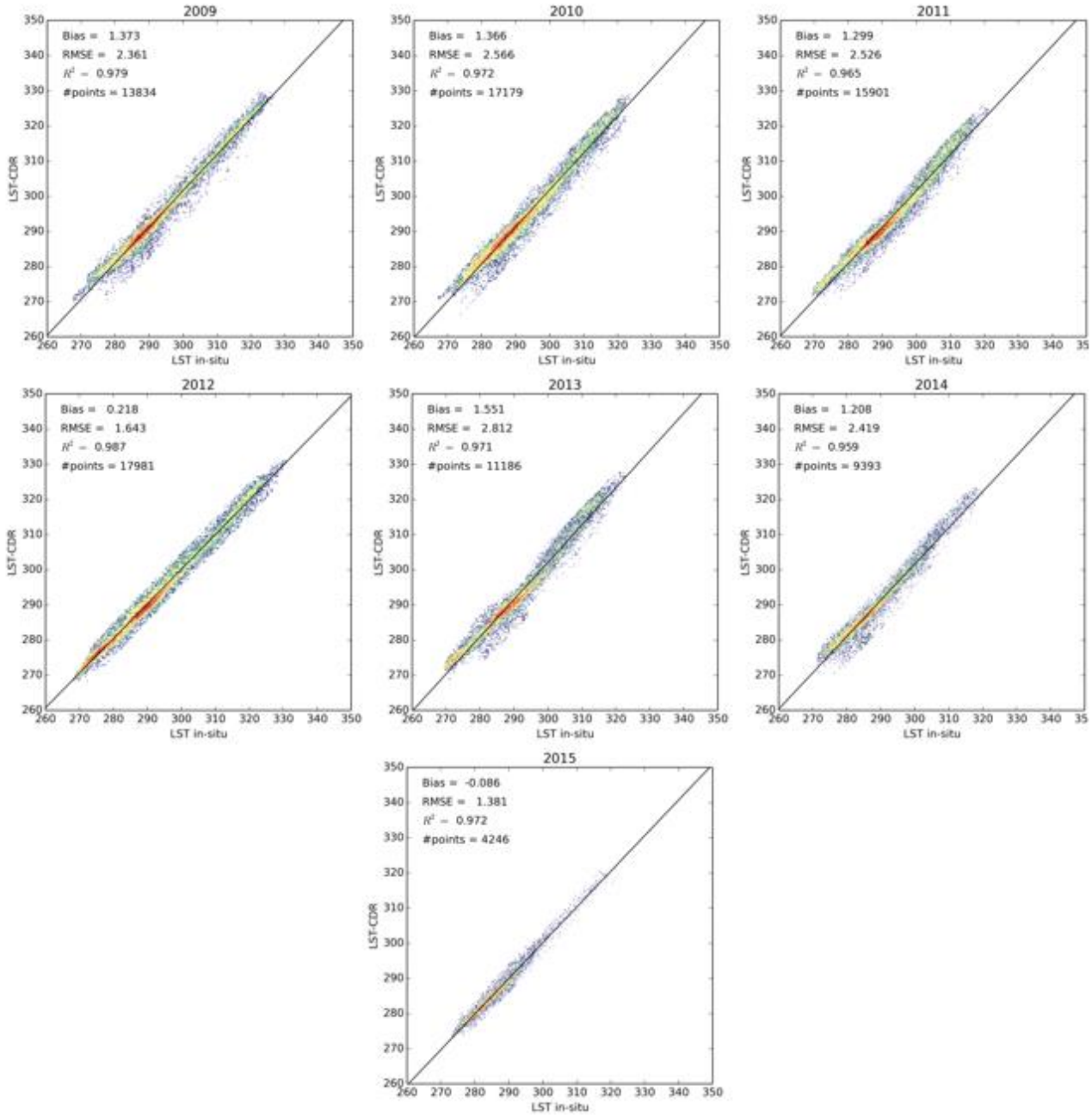


Figure 7 As in Figure 6 but respecting to Gobabeb.

The results for precision (as obtained using RMSE) in the case of Gobabeb (Figure 7) present values between 1.5 K an d 2.0 K for all considered years, with exception of 2014 that shows a slightly higher value (2.1 K). The results concerning to accuracy (bias) exhibit a similar feature meeting the

optimal accuracy for all the years with exception of 2014 (1.1 K), although still within the target accuracy.

The validation exercise in the case of RMZ Heimat Farm meets the optimal accuracy requirement for the 7 considered years, with cooler values for MLST-R in 2010 and 2014 (Figure 8). Concerning RMSE, the 2.0 K requirement is fulfilled in all the years, with exception of 2010 and 2015, although showing values closed to 2.0 K (2.1 and 2.2 K, respectively). It should be noted the lower number of match-ups between satellite and in-situ observations during 2014. The above mentioned results are summarised in Figure 10 (see corresponding station panels).

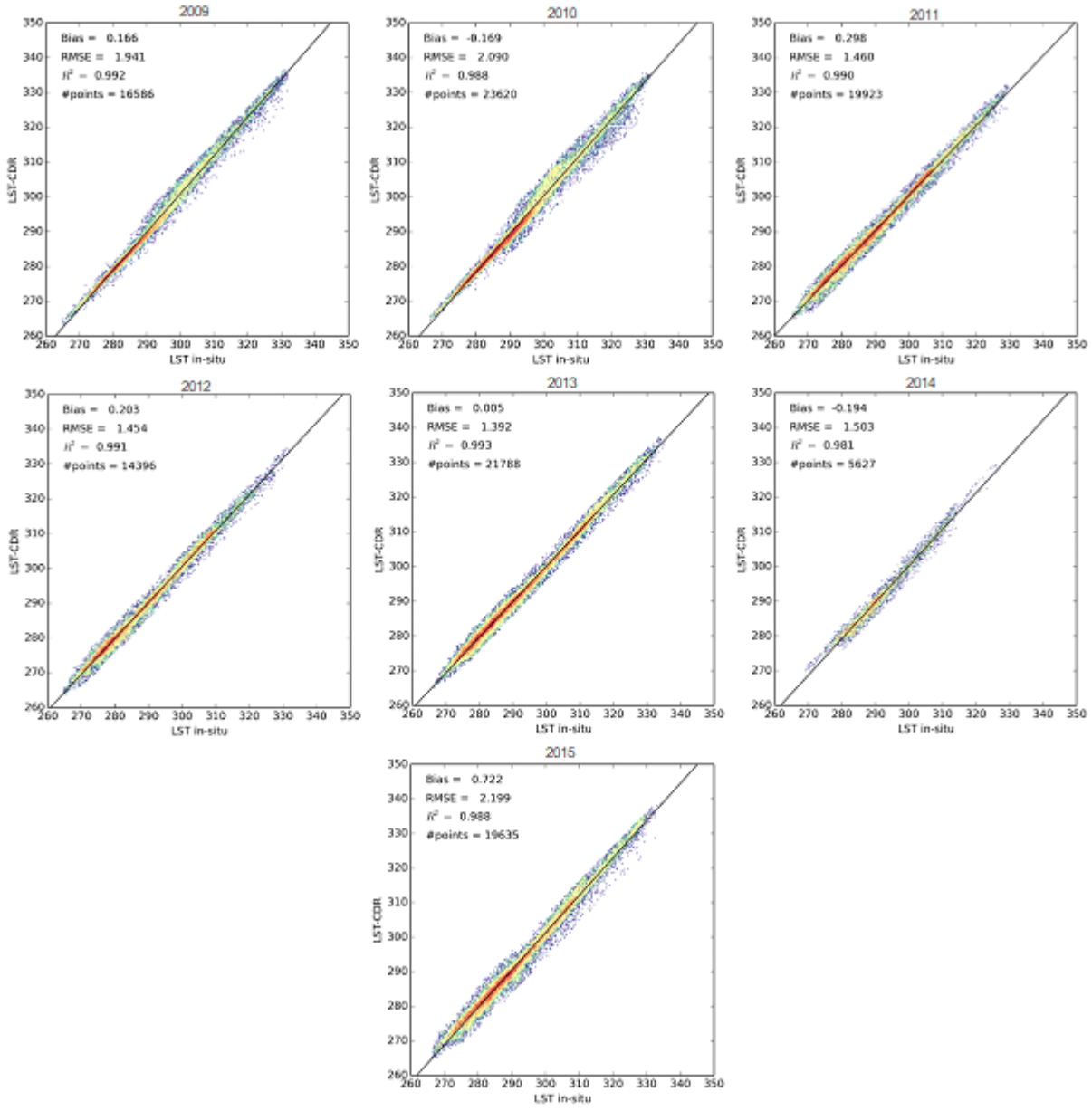


Figure 8 As in Figure 6 but respecting to RMZ Heimat Farm.

The validation exercise in the case of Dahra reveals that MLST-R meets the target accuracy requirement for all the years with exception of 2009 and 2012, when MLST-R exhibits a larger cool

bias (Figure 9). Concerning RMSE, the 2.0 K requirement is never fulfilled, although the threshold requirement of 4.0K is met for 2010, 2011, 2013 and 2014. It also should be noted the lower number of match-ups between satellite and in-situ observations in 2009, 2010 and 2012. The above mentioned results are summarised in Figure 10 (see corresponding panels). The significant cold bias, especially in 2009 and 2012, still need to further investigated, although cloud contamination, together with the occurrence of events with exceptionally high aerosol loads may be accounted for lower estimations of satellite LST when compared with in-situ observations.

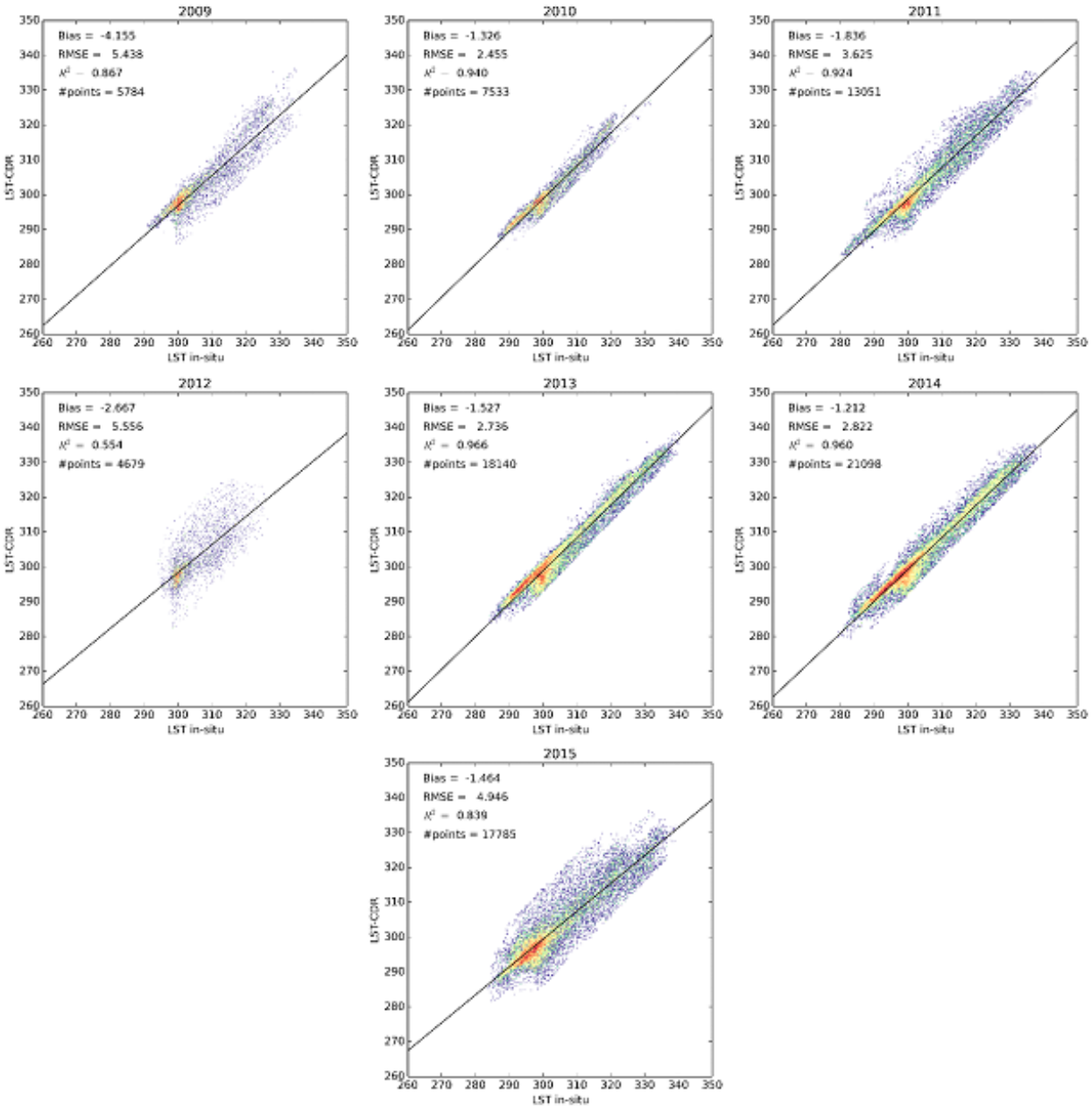


Figure 9 As in Figure 6 but respecting to Dahra.

A similar exercise was performed for daily and monthly sampling for the years from 2009 to 2012 in the sites of KIT validation stations. Figure 10 summarizes the performance measures of accuracy (RMSE in blue and Bias in orange) obtained in the validation exercise with several years of

in-situ data (2009 to 2015) for instantaneous, daily and monthly match-ups between satellite and in-situ observations for Evora, Gobabeb, RMZ Heimat and Dahra.

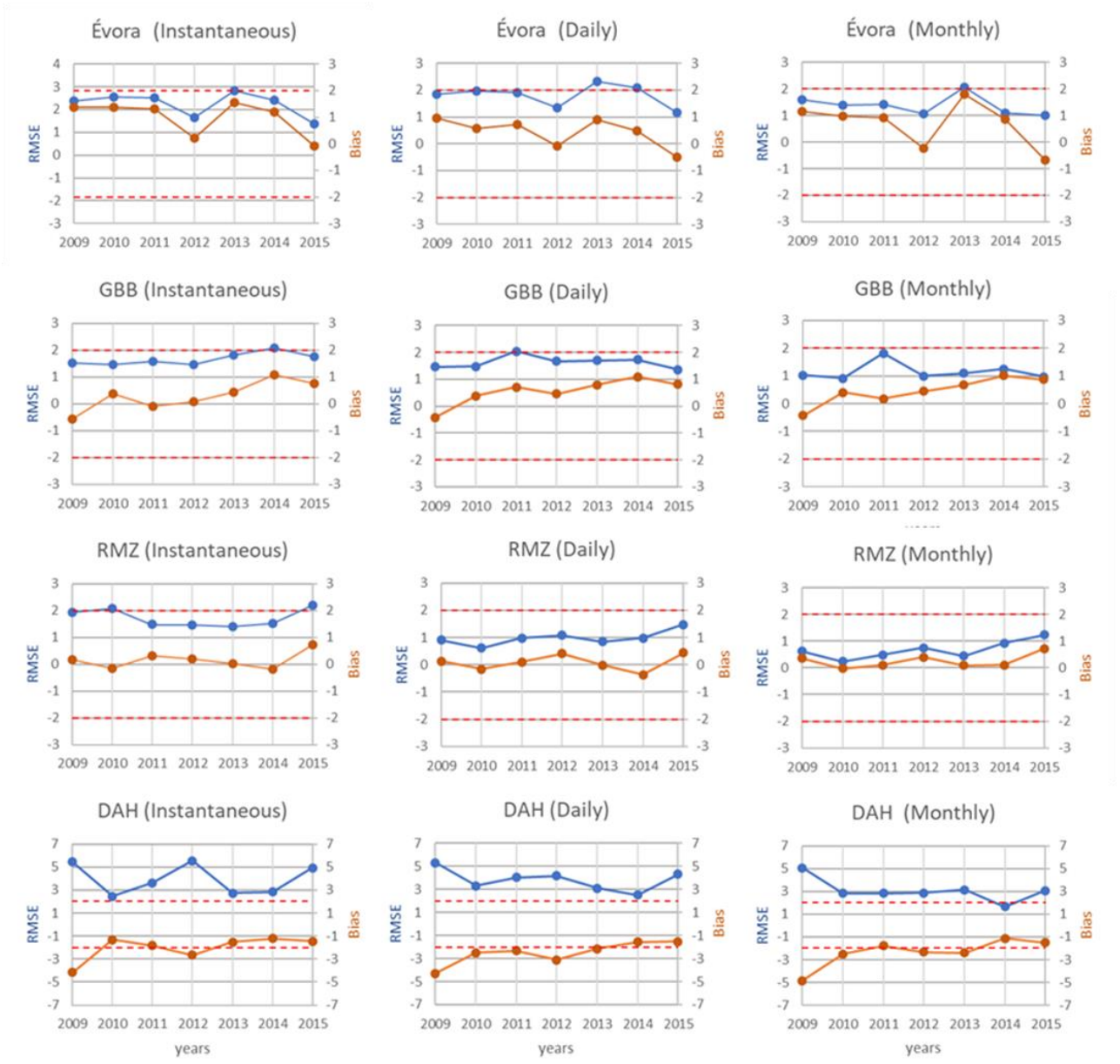


Figure 10 Instantaneous (left panel), daily (middle panel) and monthly (right panel) statistics at Evora (top panel), Gobabeb (middle top panel), RMZ Heimat (middle bottom panel) and Dahra (bottom panel). RMSE values (blue lines) refer to the left y-axis and Bias (orange lines) to the right y-axis number.

Respecting to the validation results at daily scale for Evora, Gobabeb and RMZ the obtained values of accuracy (bias) matches the optimal threshold of 1K (except Gobabeb in 2014), despite the obtained values being very close to 1 K. Additionally, RMSE values are always below 2 K, being approximately equal to 2 K for 2013 in Evora and 2011 in Gobabeb (2.1 and 2.0 K, respectively).

Dahra fulfill the 2.0 K accuracy requirement in 2014 and 2015 and the 4.0 K requirement in all years with exception of 2009.

Monthly analysis of the valid match-ups between satellite and in-situ data for the 3 sites (exception for Dahra) show a bias always lower than 1 K, with exception of Evora in 2009. The values of RMSE are all below the 2 K target accuracy. Dahra shows bias below 2.0 K during 2011, 2014 and 2015 and below 4.0 K for all years with exception of 2009.

Evora in situ LST was already compared with LSA-SAF LST NRT product (LSA-001) suggesting overall very good agreement with bias between 1 and 2K (Freitas et al., 2010, Ermida et a., 2014).

The same was also performed for Gobabeb, by comparing in-situ LST with LSA SAF NRT LST(LSA-001) for the nearest SEVIRI/MSG pixel for the period between May 2008 and March 2009 (Freitas et al., 2010). Again the agreement between the two datasets was very good, with root mean square differences between 1 and 2K. A negative bias observed in Gobabeb was already reported in 2009 in (SAF/LAND/IM/VR_LST/1.5). According to Eckardt et al. (2013) the negative bias after two rainy seasons is more marked than in other years. This may be related with the effect of grass in Gobabeb (fast desiccation) on surface emissivity, particularly the different impact within SEVIRI channel 9 and the KT15.85 IIP band. A higher dry grass fraction affects the spectrally narrower SEVIRI channel 9 less and may even contribute to reduce its emissivity, since the SEVIRI channel 9 emissivity of the background gravel & sand may be slightly higher.

Results for Dahra are in accordance with previous works (Duguay-Tetzlaff et al., 2015 and Göttsche et al., 2016) that obtained similar accuracy values for moist atmospheres mainly during rainy seasons.

In summary the optimal accuracy of 1 K is achieved for the 7 considered years in the 2 stations (Gobabeb and KMZ) and for instantaneous, daily and monthly match-ups, with the exception of 2014 in Gobabeb. In terms of RMSE, the values are below 2 K for the all cases in the 2 stations, except of Gobabeb in 2014 (instantaneous) and 2011 (daily). However, in all these exceptions the values do not exceed 2.2 K. Evora fulfils the target accuracy of 2.0 K for all years and cases and in terms of RMSE the values are close to 2.0 K. However, if only night-time instantaneous match-ups are considered, then the optimal requirement is also achieved for Evora (the station where spatial heterogeneity tends to be the largest during daytime). Dahra fulfill the 2.0 K accuracy requirement in 2014 and 2015 and the 4.0 K requirement in all years, except for 2009.

4 Temporal Stability

Decadal stability assesses the change of LST accuracy over time (i.e. per decade). Given the lack of long and continuous time-series of in situ data, and also to increase spatial representativeness, this assessment is performed by a trend analysis in bias of MLST-R time series against: i) ECMWF ERA-Interim skin temperature and ii) MODIS LST. This assessment is made for the whole time period, between 2004 and 2015.

The trend in bias is evaluated in the anomaly space, i.e. on seasonally corrected monthly MLST-R and reference data. Monthly anomalies of a given month are defined as departures from the median of that month (computed over the considered period 2004–2015). For the trend analysis, we use linear trends derived using Theil-Sen estimates (Theil 1950). The confidence intervals of the trend are estimated with the Mann-Kendall (Kendall 1938, Mann 1945).

A set of 10º longitude x 10º latitude areas are chosen to: (i) include a wide range of biomes and different atmospheric conditions within the MSG disk; (ii) include in situ stations used to validate LST_SEVIRI (Figure 1). The comparison between the MLST-R product and reference data is performed for point centred in the above mentioned areas.

Table 4 shows the acronyms of the six selected areas, as well as the latitude and longitude of the centred point and Figure 12 shows the main land cover types in the areas of 10º longitude x 10º latitude, as obtained from the IGBP database (Belward, 1996); the selected smaller boxes of 0.75º x 0.75º (red squares) centred within the former are also shown. The stability assessment will be performed on the data available over the selected boxes of 0.75º x 0.75º.

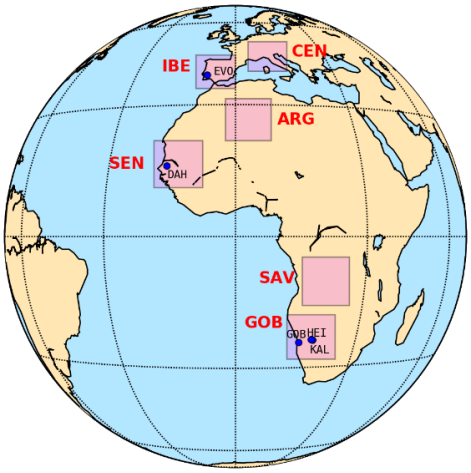


Figure 11 Areas (10º longitude x10º latitude) considered for comparison between MLST-R and reference data: IBE, CEN, ARG, SAV, SEN and GOB.

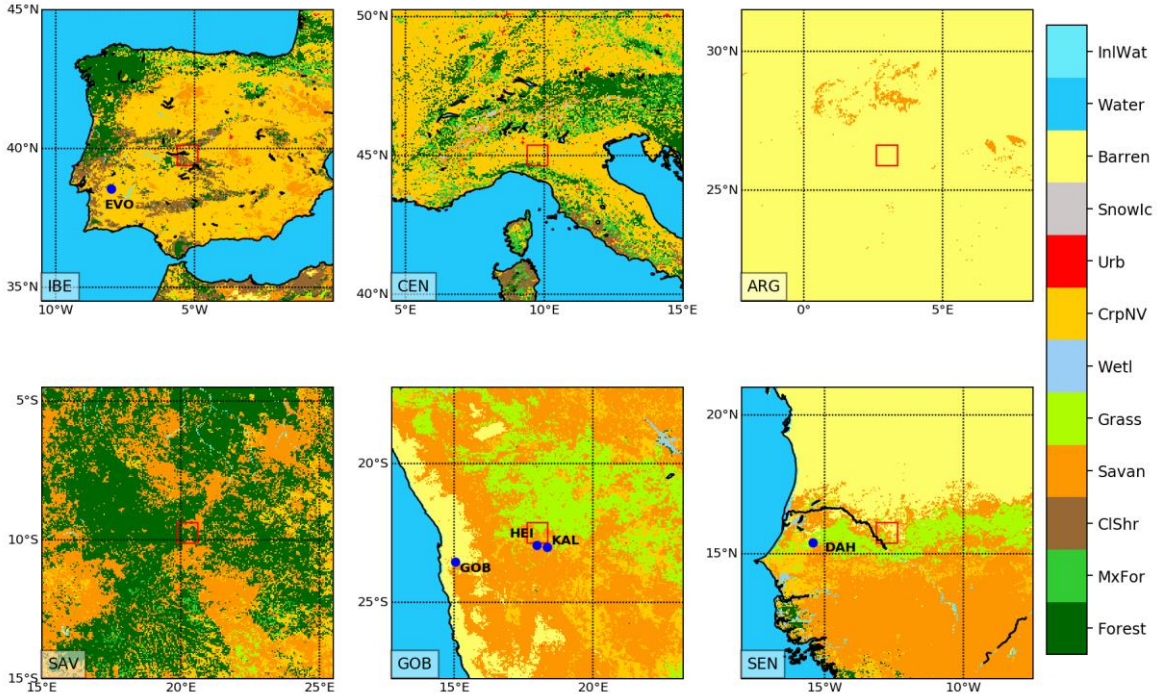


Figure 12 Land cover types as obtained from IGBP database (Belward, 1996) for the six defined areas. Red squares represented the selected boxes of 0.75º x 0.75 º centred in each area used to The stability assessment

Table 4 Acronym of the six selected areas (10º longitude x10º latitude) and the latitude and longitude of the centred point.

Name	Iberian Peninsula	Central Europe	Algeria	Senegal	Southern Africa/ Savannah	Namibia
Acronym	IBE*	CEN	ARG#	SEN	SAV	GOB$
Center [lat, lon]	[5ºW,40ºN]	[10ºE,45ºN]	[3ºE,26ºN]	[12.5ºE,15.5ºN]	[20ºE,10ºS]	[18ºE,23.5ºS]

* includes Evora station; # includes Dahra station; $ includes Gobabeb and Farm Heimat stations.

4.1 Comparison between MLST-R and ECMWF Skin Temperature

The first step to evaluate temporal stability of MLST-R is made using the skin temperature (T_{skin}) obtained from ERA-Interim reanalysis produced by ECMWF. Data from ERA-Interim reanalysis are available since 1979 on a grid of approximately 0.75º spatial resolution, four times a day (00, 06, 12, and 18 UTC). ECMWF model T_{skin} represents the temperature of the interface between soil and atmosphere; it is an instantaneous variable with no memory. Therefore, it is comparable to the land surface temperature observed by satellites, apart from the directional effects that affect the latter. Moreover, it should be noted that MLST-R retrieval uses total column water vapour from ERA-Interim, but this is only considered for the choice of the suitable set of generalized split-window coefficients for each LST retrieval. As such, ERA-Interim water vapour column is used implicitly (but not explicitly) by the MLST-R CDR.

ERA-Interim skin temperatures (ECMWF-Tskin) is extracted for the six areas mentioned before (Figure 12) and the ECMWF-Tskin were re-sampled in space and time to match the MLST-R temporal and spatial resolution. The decadal stability of MLST-R evaluated against the ECMWF-Tskin is then assessed over a 0.75º x 0.75º grid box located at the center of each of the six considered validation areas (see Figure 11 and Figure 12), for 00, 06, 12 and 18 UTC (i.e., the analyses verification times). The seasonal cycles of the valid points of MLST-R and ECMWF-Tskin obtained after match-up were obtained for the six areas (Figure 11). MLST-R is considered cloudy if more than 50% of the valid points within the ECMWF grid point are classified as cloudy, whereas ECMWF-T_{skin} is considered as cloudy if the ECMWF Total Cloud Cover (TCC) is higher than 0.3.

Figure 13 shows a marked seasonal cycle of MLST-R and ECMWF-T_{skin} over Algeria, Central Europe, Iberia and Gobabeb (Namibia) and a less marked seasonal cycle over Savana and Senegal, which may be associated with the main land cover types present in the 0.75º x 0.75º grid centered areas (see Figure 10), namely bare soil and forest. While Algeria, Central Europe, Iberia and Gobabeb show the minimum temperature values during the Northern Hemisphere Winter (DJF), Savana and Namibia exhibit the minimum during the Southern Hemisphere winter (JJA). High seasonal variability is observed for Senegal and Iberia boxes. It should also be mentioned that the vegetation high seasonality in those regions, where ground (inter-canopy) vegetation in sparsely-to-moderately vegetated areas in Senegal and Iberia becomes desiccated during the dry season, leads to large spatial heterogeneity, particularly during daytime (see e.g., Ermida et al., 2014). In the 0.75º x 0.75º grid centered in Iberia, MLST-R shows higher values than ECMWF-T_{skin}, namely in Summer and Winter, whereas in Senegal the MLST-R annual maximum values are lower than ECMWF-T_{skin}. The

number of valid points are usually lower in wet months (more clouds), being lower in case of MLST-R. Over Savana region the number of valid pixels is very low and even null during the wet season, namely in the case of ECMWF-T_{skin}, and several monthly values were removed from the trend analysis Figure 13).

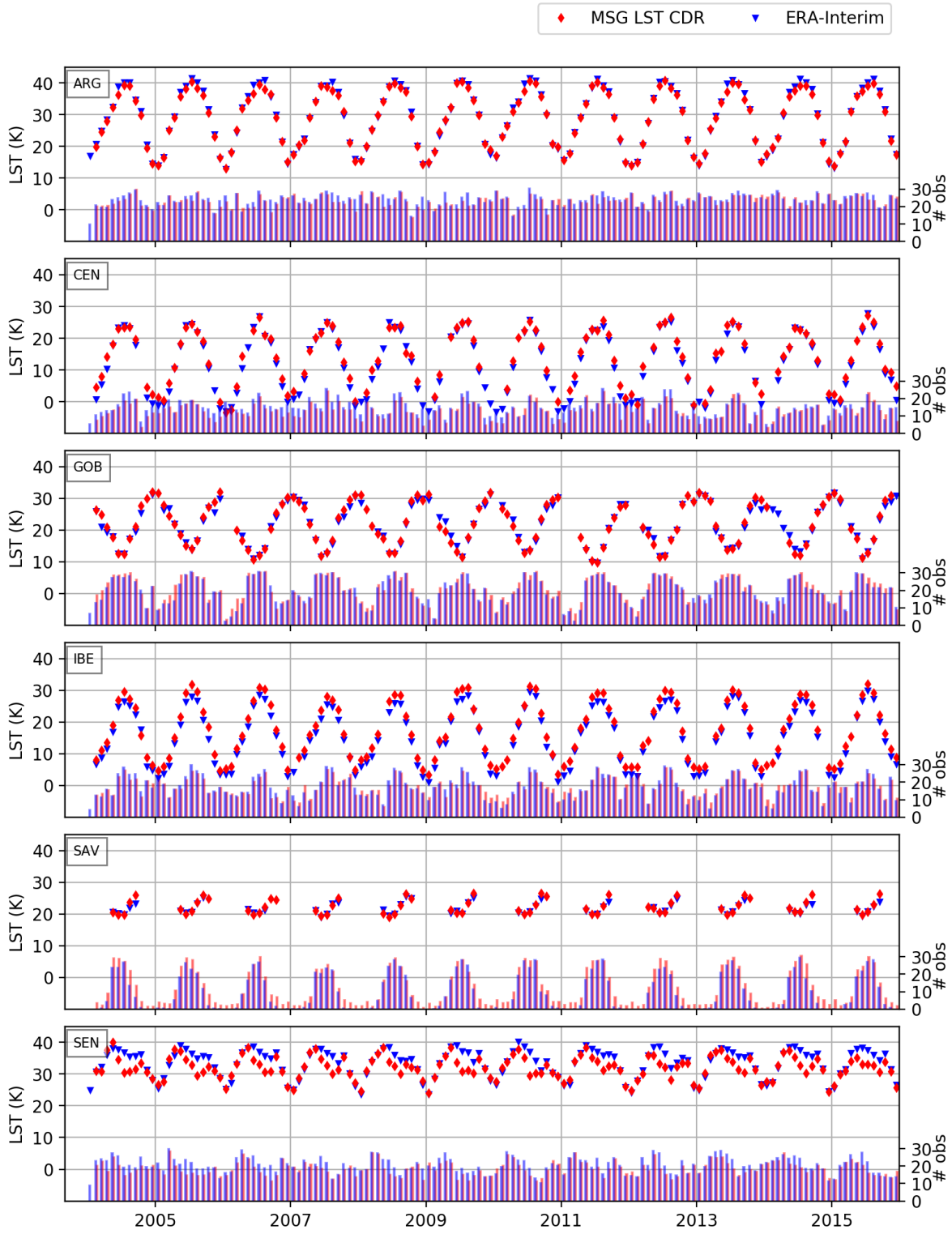


Figure 13 Time series of monthly mean values of MLST-R and ECMWF-T_{skin} for the 0.75º x 0.75º grid box located at the centre of each of the six considered areas (left y-axis) and the number of valid points obtained by the match-up between the 2 datasets (right y-axis).

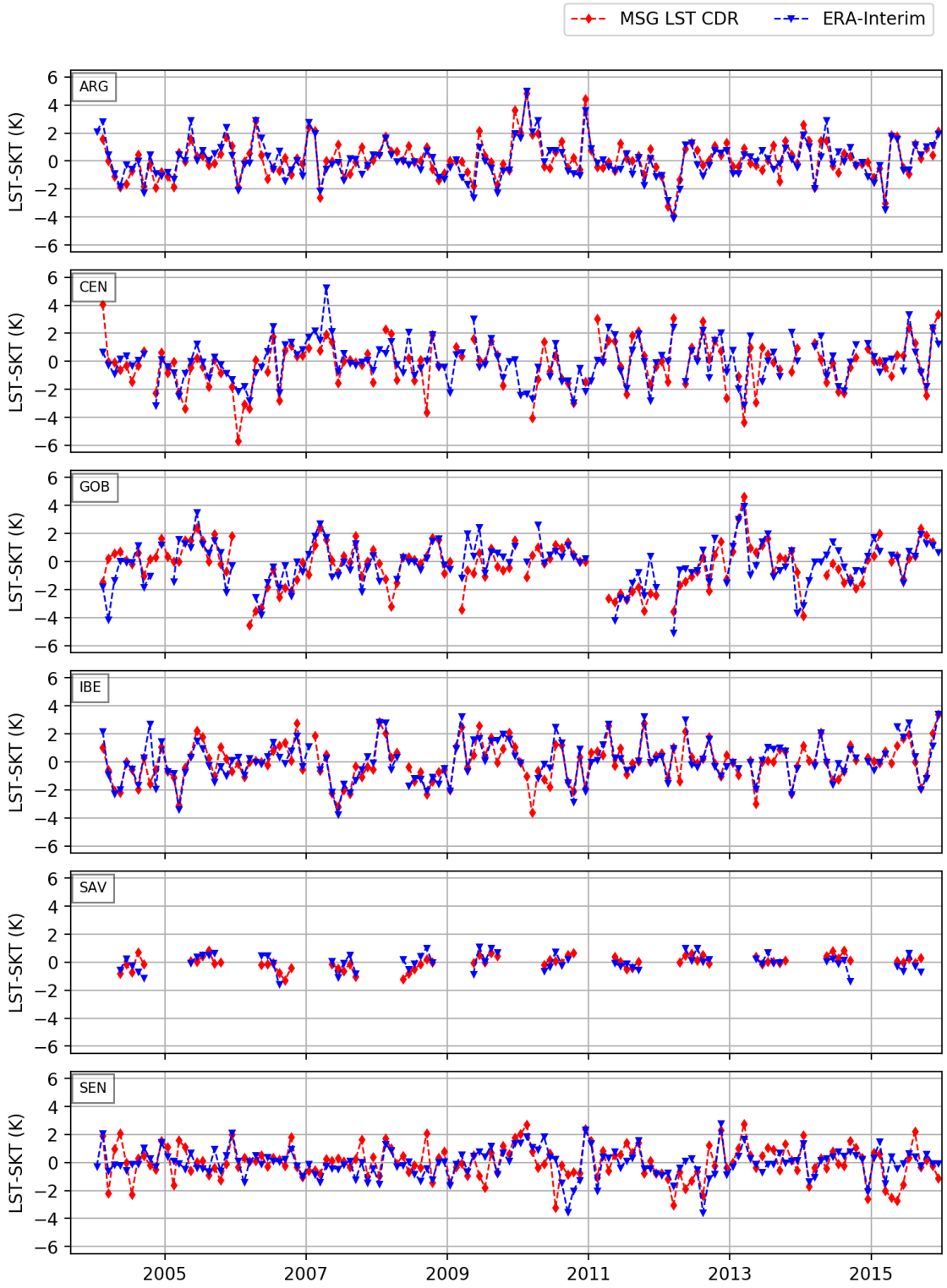


Figure 14 Time series of monthly anomalies of averaged at a 0.75° x 0.75° latitude and longitude window centred at the 6 considered areas for the period from 2004 to 2015.

The time series of MLST-R and ECMWF-T_{skin} monthly anomalies for the 0.75º x 0.75º latitude and longitude window centered at the six considered areas (Figure 14) show an overall good agreement between the two time-series. The higher differences were observed over Namibia (GOB)

with higher (lower) values of MLST-R anomalies when compared with ECMWF-T_{skin} during the initial (final) part of the analyzed period. Isolated high differences are also observed during the last years of the considered period in Central Europe and Senegal.

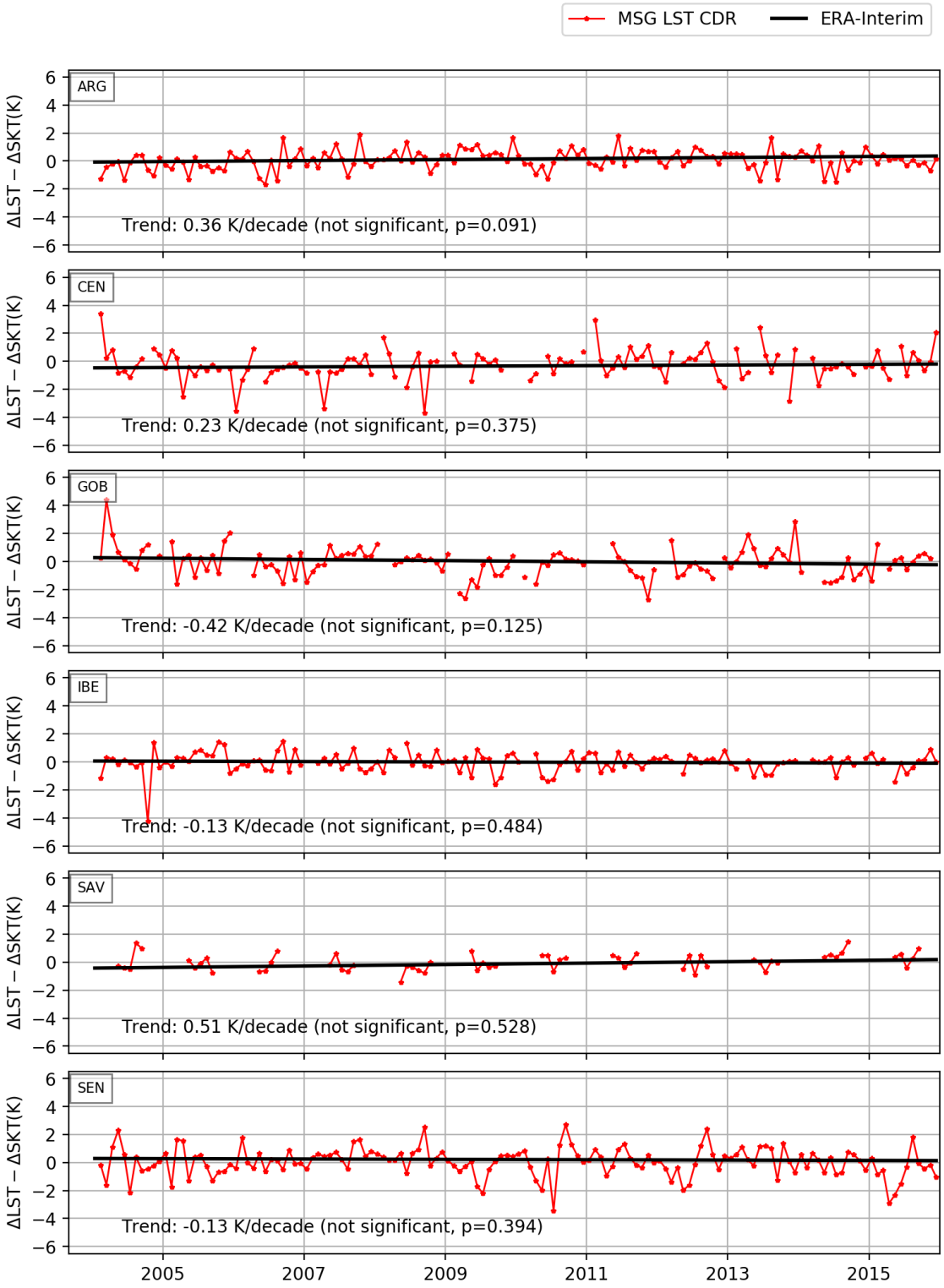


Figure 15 Time series of mean bias between MLST-R and ECMWF-Tskin at a 0.75° x 0.75° latitude and longitude window centred at the 6 considered areas for the period from 2004 to 2012. The black line represents the Theil-Sen linear trend.

The difference between the temperature anomalies in Figure 14 (MLST-R – ECMWF T_{skin}) is shown for each analyzed grid cell in Figure 15, and the trend in the difference as well as its significance and p-value, are indicated within each plot. Non-statistically significant (above the 5% significance level) positive decadal trends in bias between MLST-R and ECMWF-T_{skin} are observed over Algeria, Central Europe and Savana (0.36, 0.23 and 0.51 K/decade, respectively), whereas a non-significant negative decadal trend is observed over Namibia, Iberia and Senegal with values of -0.42, -0.13 and -0.13 K/decade, respectively (Figure 13). It should be noted that the bias between the two time-series in Senegal present high values, namely in the early and later years of the reprocessed period. The decadal trend of bias over Savana area does not present a significant trend, despite the few number of monthly values available. This makes, however, the trend assessment very challenging and less representative for such a short period.

The 0.2 K stability requirement is achieved at two (Senegal and Iberia) out of six investigated locations. At Central Europe and Algeria boxes, the observed decadal trend is within 0.4 K, while Gobabeb and Savana exhibit higher decadal trend values with opposite directions (positive for Savana and negative for Gobabeb), although not statistically significant.

In general, the comparison presents features, which are similar to those found in previous comparisons between model skin temperatures and satellite LST (e.g., Trigo et al., 2015; Wang et al., 2014; Zheng et al., 2012; Orth et al., 2017). These include the tendency for model surface temperatures to have a lower daily amplitude particularly over arid and semi-arid regions during the dry season.

4.2 Comparison between MLST-R and MODIS LST

MODIS is one of the most advanced passive imaging instrument flying on board Terra and Aqua polar-orbiting satellites, acquiring information in 36 spectral bands. Both orbits are sun synchronous and Terra's orbit passes from north to south across the equator in the morning (10:30, local solar time), whereas Aqua passes south to north in the afternoon (1:30, local solar time). Terra MODIS and Aqua MODIS revisiting times are about 1 to 2 days. Although there are some well-known problems over some MODIS bands, each collection/version of their products mitigates those problems with newer calibration coefficients (e.g. Madhavan et al, 2016).

Level-2 v006 MOD11_L2 and MYD11_L2 from Terra and Aqua (respectively) are used for the whole period of the MLST-R dataset (2004-2015) over the 6 areas of Figure 11. The swath data are then re-projected with a nearest neighbour technique to the SEVIRI grid, averaging the valid MODIS values within a given MLST-R pixel. The matchup with the corresponding MLST-R values are performed in time using the nearest SEVIRI value in time if it contained a valid value (i.e. collocation is within 7.5 min at most), taking into account the acquisition time of both sensors. The high frequency of MLST-R product (15 minutes) increases the chances of getting good matchups between the two satellites. In contrast, collocation in space and time between polar-orbiter observations is much more difficult, and not included in this report.

The accuracy of MODIS LST products, as obtained through validation field campaigns and radiance-based studies, is generally better than 1 K (Wan et al. 2002). Arid regions exhibit the higher errors due to uncertainties in emissivity values and heavy dust aerosols. The contamination by

clouds and heavy aerosols may lead to errors from 4 to 11K (Wan et al. 2002). The latest version of the product (used in this validation exercise), mitigates some of these deficiencies (Wan, 2014).

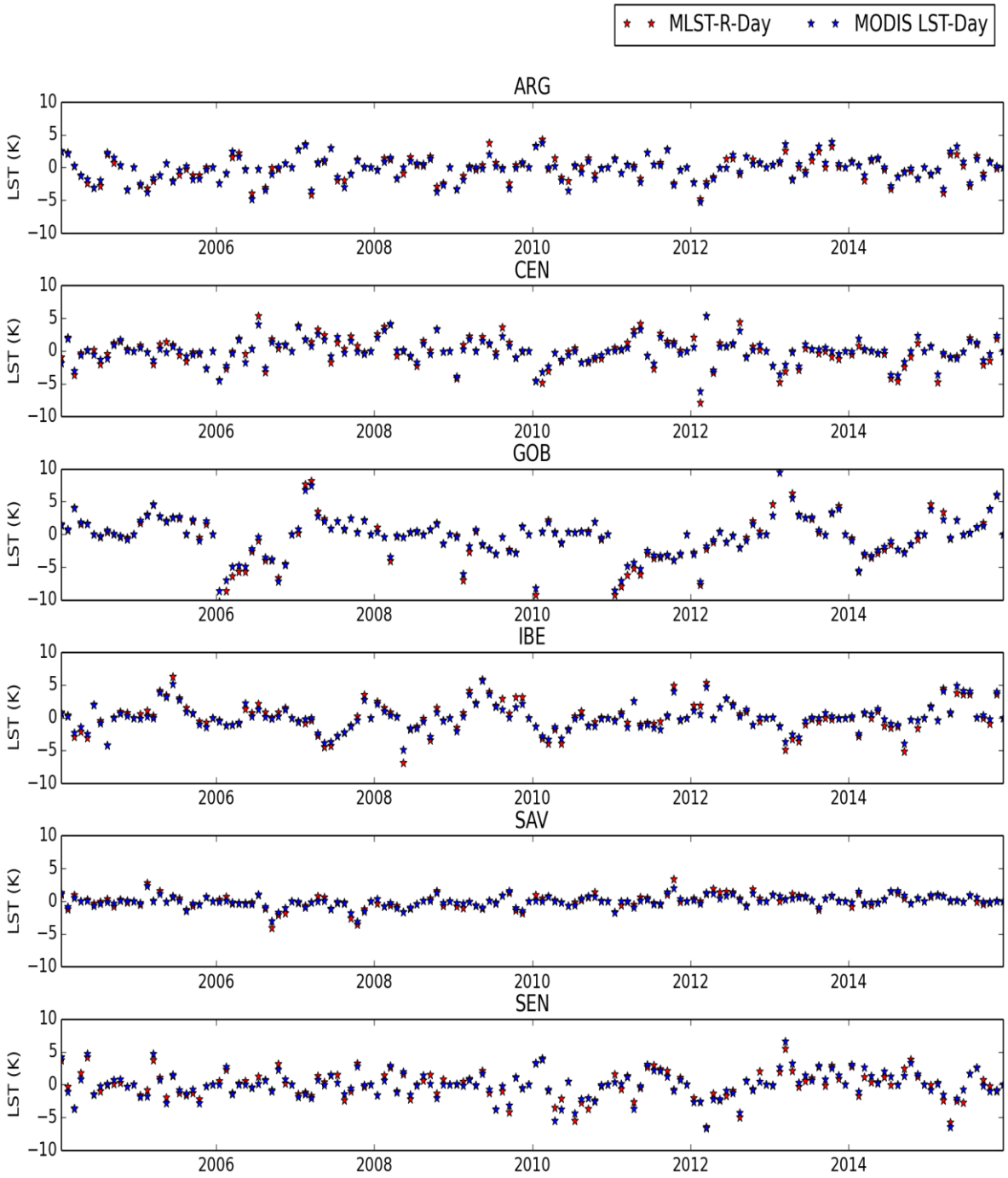


Figure 16 Time series of monthly anomalies of MLST-R and LST-MODIS averaged at a 1° x 1° latitude and longitude window centred at the 6 considered areas for the period from 2004 to 2015 considering day time match-ups.

The decadal stability of the MLST-R is now evaluated against equivalent MODIS product for the period 2004 to 2015 over a 1° x 1° latitude and longitude window centred at the six defined

areas (Figure 16 to Figure 19). We have used both MODIS day- and night-time LSTs. Monthly anomalies are computed over a 1° x 1° latitude and longitude window centred at the six defined areas for both day- and night-time (Figure 16 and Figure 17). Higher anomalies were obtained, as expected, for day-time match-ups and for Algeria, Iberia, Gobabeb and Senegal. High positive (negative) anomalies were obtained in Gobabeb for 2007, 2013 and 2015 (2006 and 2011) for day time match-ups. The 1º x 1º window centred in Savana region presents the lowest anomaly values for day and night- time (Figure 17). It should be noted that the anomalies from both datasets are very well correlated, despite they are measured by different sensors and using slightly different algorithms and input data; this emphasizes the environmental character of the factors causing the anomaly, which therefore affect both (observational) datasets in a similar way. This was not so clear in the comparisons with ERA-Interim, since these environmental constraints affect the model (and its scale) in a different way.

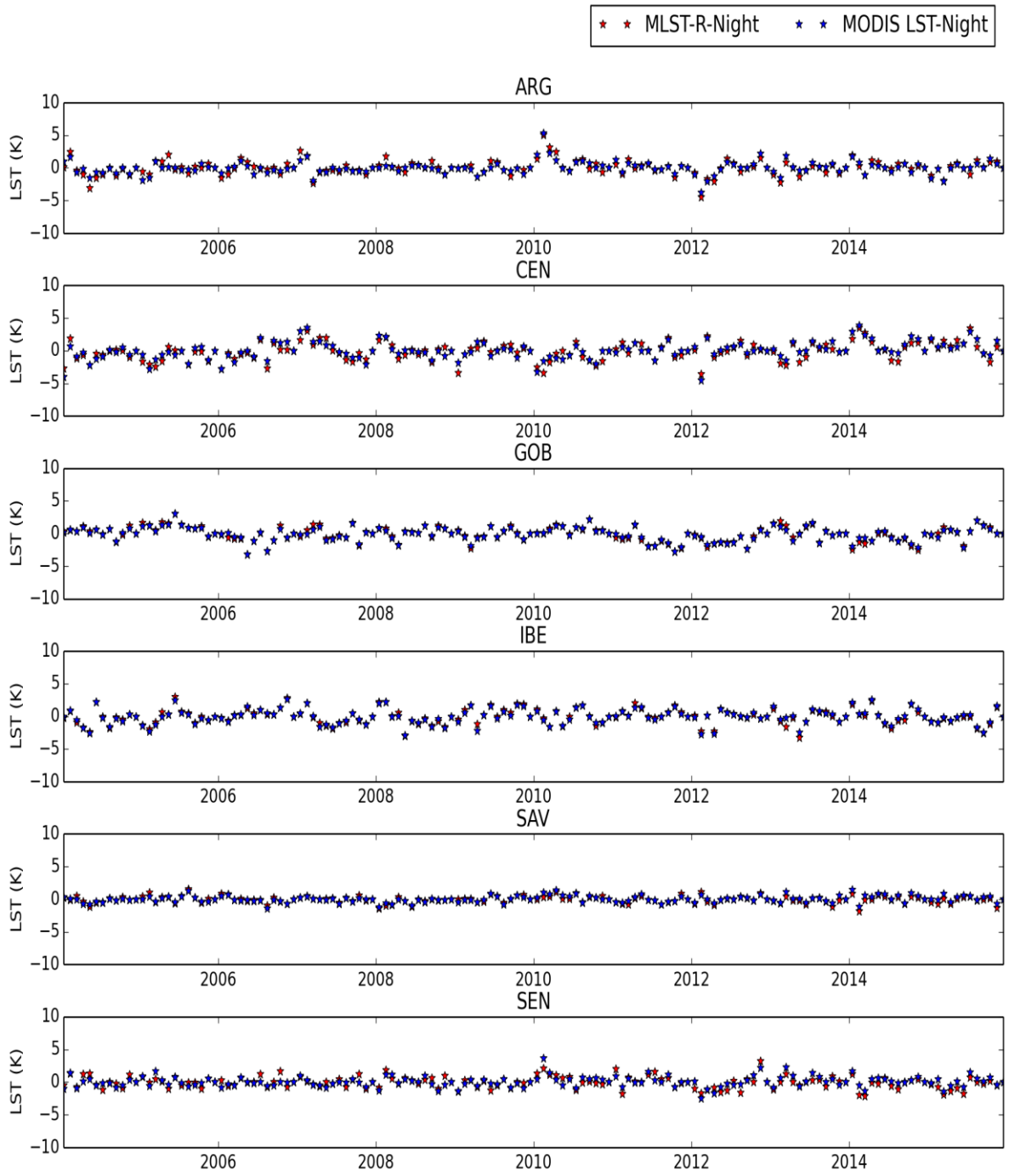


Figure 17 As in Figure 16 but considering night time match-ups.

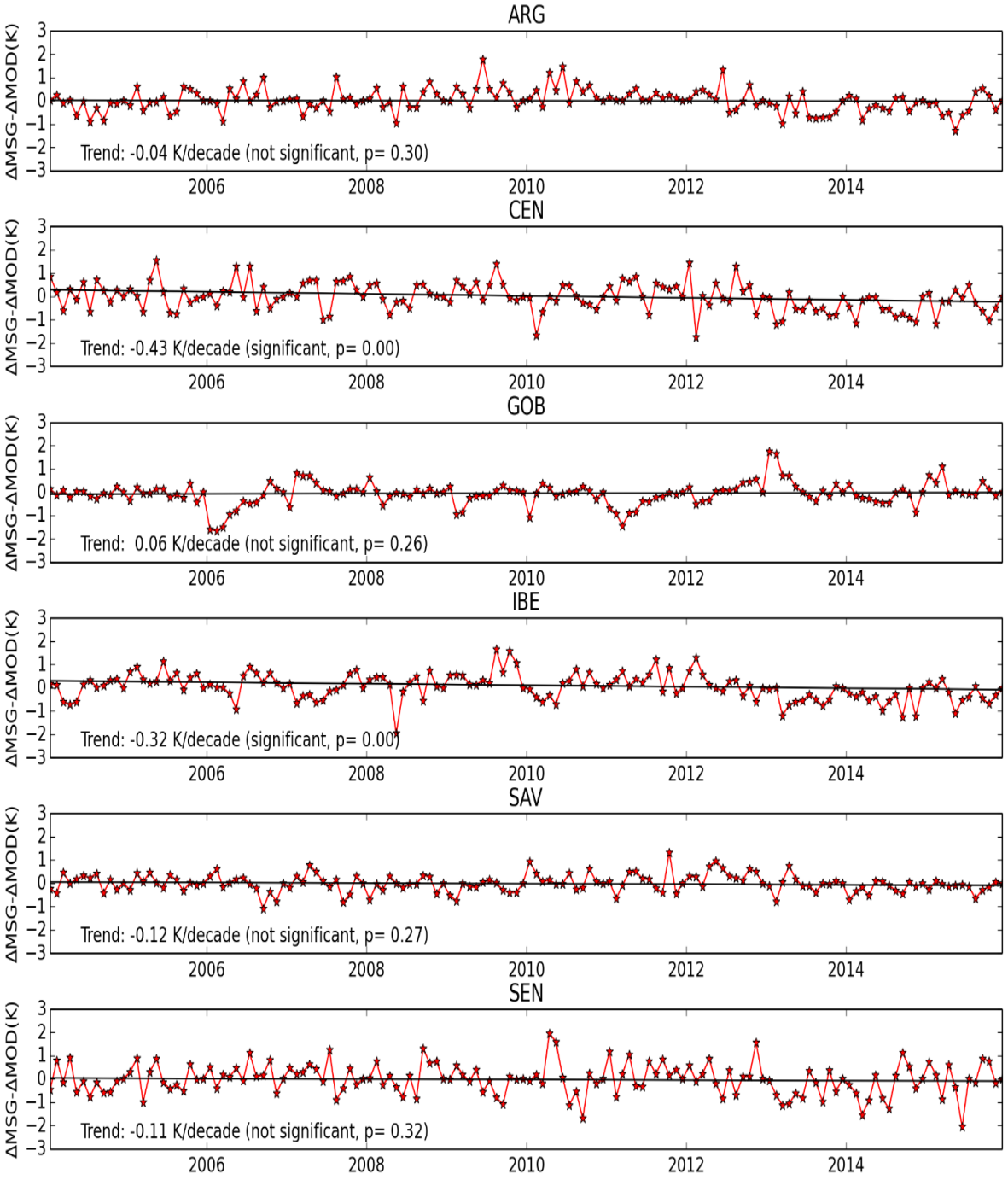


Figure 18 Time series of mean bias between MLST-R and LST-MODIS at a 1° x 1° latitude and longitude window centred at the 6 considered areas for the period from 2004 to 2015 considering day time match-ups. The black line represents the Theil-Sen linear trend.

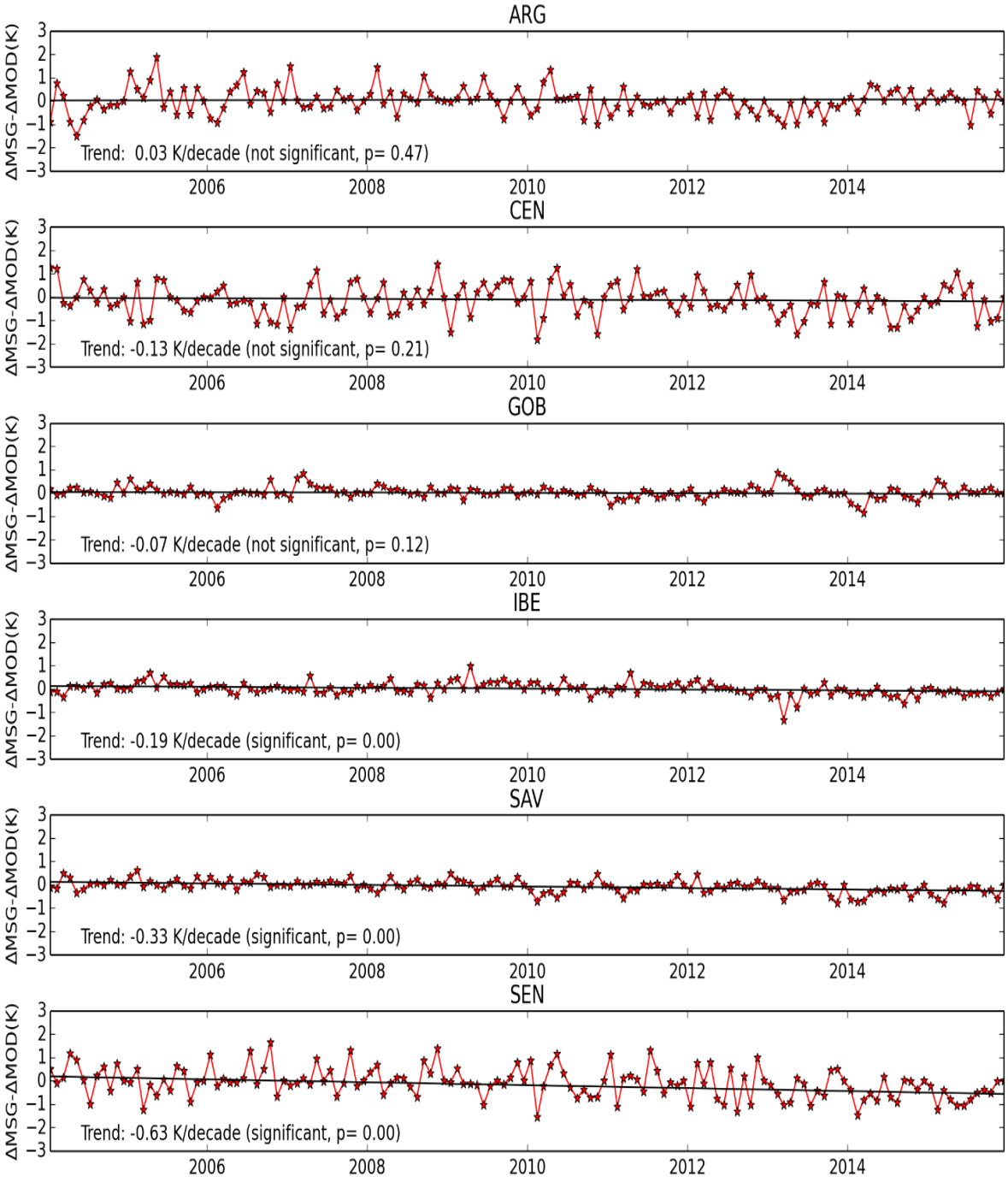


Figure 19 As in Figure 18, but considering night time match-ups.

This MODIS-based stability exercise is shown in Figure 18 and Figure 19 for all investigated regions for both day and night match-ups. The stability over Savana and Iberia is below 0.4 K/decade, respectively, for day and night-time. A better (within 0.2 K/decade) stability is reached four at out of six investigated locations for both day- and night-time match-ups (Algeria, Gobabeb, Savana, Senegal for day-time and Algeria, Central Europe, Gobabeb and Iberia for night-time match-ups). Furthermore, the decadal trends obtained were not significant (significance level above 5%) over Algeria, Gobabeb, Savana and Senegal for day-time match-ups and over Algeria, Central Europe and Gobabeb for night-time match-ups, being negative for all the regions, with exception of

Gobabeb for day-time and Algeria for night-time match-ups. Instrument drifts or the short time-span of the dataset are likely the cause of these trends, but more work would be required to determine their exact causes.

5 Conclusions

This report aims to document the validation of MLST-R (LSA-050) product. Reference datasets from independent observation sources, reanalysis and from independent satellite measurements were used. Concerning the in-situ validation, four permanent LST validation stations operated by Karlsruhe Institute of Technology (KIT) are used. They are located in areas characterized by naturally homogeneous land use and land cover in different climate zones.

In-situ validation exercises performed over Evora, Gobabeb and RMZ show the fulfilment of the optimal accuracy of 1 K is achieved for the 7 considered years in the 2 stations (Gobabeb and EMZ) for instantaneous, daily and monthly match-ups, except 2014 in Gobabeb. In terms of RMSE, the values are below 2 K for the all cases in the 2 stations, except for Gobabeb in 2014 (instantaneous) and 2011 (daily). Nevertheless, RMSE never exceed 2.2 K. Evora fulfils the target accuracy of 2.0 K for all years and cases and in terms of RMSE the values are close to 2.0 K. It should be stressed, however, that if only night-time instantaneous match-ups are considered, the optimal requirement is also achieved for Evora. This is in agreement with the characteristics of the surrounding station terrain, where spatial heterogeneity tends to be largest during daytime. Dahra fulfills the 2.0 K accuracy requirement in 2014 and 2015 and the 4.0 K requirement in all years with exception of 2009. It is acknowledged that further investigation related with accuracy and precision should be done including other regions, such as mountain areas or areas with high viewing angles, where this validation exercise could not be made due to the lack of in-situ data.

The decadal stability is performed by means of computing the bias between MLST-R and the ECMWF ERA-Interim skin temperature for a window of 0.75º x 0.75º centred in the above mentioned areas. Results indicate that a trend of 0.2 K/decade is reached, suggesting the fulfillment of 0.2K/decade stability requirement at two (Senegal and Iberia) out of six investigated locations. A stability of up to 0.4 K/decade was observed in Central Europe and Algeria boxes. On the other hand, Gobabeb and Savana exhibit higher decadal trend values with opposite directions (positive for Savana and negative for Gobabeb), although not significant.

The decadal stability of the MLST-R was also evaluated against the Level-2 v006 MOD11_L2 and MYD11_L2 from Terra and Aqua, over a 1° x 1° latitude and longitude window centred at the six defined areas. This MODIS-based stability revealed valued below 0.8 K/decade over all investigated regions, for both day and night-time match-ups (Figure 16 and Figure 17). A stability of 0.4 K/decade is observed in Savana and Iberia respectively for day and night-time, and better performance, with 0.2 K/decade or less, is reached at out of six investigated locations for both day- and night-time match-ups (Algeria, Gobabeb, Savana, Senegal for day-time and Algeria, Central Europe and Gobabeb for night-time match-ups). Furthermore, the decadal trends obtained were not significant (significance levels above 5%) over Algeria, Iberia, Savana and Senegal for day-time match-ups and over Algeria, Central Europe and Iberia for night-time match-ups, being negative for all the regions, with exception of Gobabeb for day-time and Algeria for night-time match-ups.

6 References

Belward AS. 1996. The IGBP-DIS global 1km land cover data set (DISCover) – proposal and implementation plans. Toulouse, France.

Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., Adler-Golden, S. M., Chetwynd Jr., J. H., Richtsmeier, S. C., Pukall, B., Allred, C. L., Jeong, L. S.; Hoke, M. L. (1999). MODTRAN4 radiative transfer modeling for atmospheric correction. Proc. SPIE, 3756, 348–353.

Eckardt, F.D., Soderberg, K.L., Coop, J., Muller, A.A., Vickery, K.J., Grandin, R.D., Jack, C., Kapalanga, T.S., and Henschel, J. (2013). The nature of moisture at Gobabeb, in the central Namib Desert. Journal of Arid Environments, Vol. 93.

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 Sens. Environ., Vol. 148, pp. 16-27.

Freitas, S. C., I. F. Trigo, J. M. Bioucas-Dias, and F. Göttsche, 2010, Quantifying the Uncertainty of Land Surface Temperature Retrievals from SEVIRI/METEOSAT. IEEE Trans. Geosci. Remote Sens., Vol. 48, Num. 1, pp. 523-534.

French, A.N., Schmugge, T. J., and Kustas, W.P. (2000). Discrimination of Senescent Vegetation Using Thermal Emissivity Contrast. Remote Sensing of Environment, Vol. 74, pp. 249–254.

Göttsche, F.-M., and Hulley, G.C. (2012). Validation of six satellite-retrieved land surface emissivity products over two land cover types in a hyper-arid region. Remote Sens. Environ., Vol. 124, pp. 149-158.

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. Inter. J. of Remote Sens., Vol. 34, Issue 9-10, pp. 3069-3083.

Göttsche F-M, Olesen F-S, Trigo I, Bork-Unkelbach A, Martin M. (2016). Long Term Validation of Land Surface Temperature Retrieved from MSG/SEVIRI with Continuous in-Situ Measurements in Africa. Remote Sensing. Multidisciplinary Digital Publishing Institute 8(5): 410. DOI: 10.3390/rs8050410.

Guillevic, P.C., Bork-Unkelbach, A., Göttsche, F.M., Hulley, G., Gastellu-Etchegorry, J.-P., Privette, J.L., and Olesen, F.S. (2013). Directional viewing effects on Land Surface Temperature products over sparse vegetation canopies - A multi-sensors analysis from field to polar to geostationary satellites measurements. IEEE Transactions on Geoscience and Remote Sensing, Vol. 10(6), pp. 1464–1468.

Hulley, G.C., Hook, S.J., Manning, E., Lee, S.-Y., and Fetzer, E. (2009). Validation of the Atmospheric Infrared Sounder (AIRS) version 5 land surface emissivity product over the Namib and Kalahari deserts. Journal of Geophysical Research, Vol. 114(D19).
Lancaster, J., Lancaster, N., and Seely, M.K. (1984). Climate of the central Namib Desert. Madoqua, Vol. 14(1), pp. 5–61.

Jimenez-Munoz, J.C., Sobrino, J.A., Mattar, C., Hulley, G., and Gottsche, F.-M. (2014). Temperature and emissivity separation from MSG/SEVIRI data. IEEE Transactions on Geoscience and Remote Sensing, Vol. 52(9), pp. 5937–5951.

Kendall, M. G. (1938). «A NEW MEASURE OF RANK CORRELATION». Biometrika. 30 (1-2): 81–93. ISSN 0006-3444. doi:10.1093/biomet/30.1-2.81

Madhavan, S., Xiong, X., Wu, A., Wenny, B. N., Chiang, K., Chen, N., Wang, Z., Li, Y. (2016). Noise Characterization and Performance of MODIS Thermal Emissive Bands, IEEE Transactions on Geoscience and Remote Sensing, vol. 54(6), pp. 3221-3234, doi: 10.1109/TGRS.2015.2514061

Mann, H. (1945). Nonparametric Tests Against Trend. Econometrica, 13(3), 245-259. doi:10.2307/1907187

Olioso, A., Sòria, G., Sobrino, J., and Duchemin, B. (2007). Evidence of low land surface thermal infrared emissivity in the presence of dry vegetation. Geoscience and Remote Sensing Letters, IEEE, Vol. 4(1), pp. 112–116.

Orth, R., Dutra, E. , Trigo, I.F. and Balsamo, G.: Advancing land surface model development with satellite-based Earth observations. Hydr. Earth Syst. Sci., 2017. (doi: 10.5194/hess-2016-628).

Peel, M.C., Finlayson, B.L., and McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences, 11:1633–1644, 2007.

Rubio, E., Caselles, V., and Badenas, C. (1997). Emissivity measurements of several soils and vegetation types in the 8–14 µm wave band: analysis of two field methods. Remote Sensing of Environment, Vol. 59, pp. 490–521.

Schädlich, S., Göttsche, F.-M., and Olesen, F.-S. (2001). Influence of Land Surface Parameters and Atmosphere on METEOSAT Brightness Temperatures and Generation of Land Surface Temperature Maps by Temporally and Spatially Interpolating Atmospheric Correction. Remote Sensing of Environment, Vol. 75(1), pp. 39–46.

Schneider, P., Ghent, D., Corlett, G., Prata, F., and Remedios, J. (2012). AATSR Validation: LST Validation Protocol. Report, ESA Contract Number: 19054/05/NL/F (UL-NILU-ESA-LST-LVP Issue1 Rev0), pp. 1-39.

Theil, H. (1950), "A rank-invariant method of linear and polynomial regression analysis. I, II, III", Nederl. Akad. Wetensch., Proc., 53: 386–392, 521–525, 1397–1412, MR 0036489.

Trigo, I. F., L. F. Peres, C. C. DaCamara, and S. C. Freitas (2008a), Thermal Land Surface Emissivity retrieved from SEVIRI/METEOSAT, IEEE Trans. Geosci. Remote Sens., 46, doi: 10.1109/TGRS.2007.905197.

Trigo, I. F., I. T. Monteiro, F. Olesen, and E. Kabsch (2008b), An assessment of remotely sensed Land Surface Temperature. J. Geophys Res., 113, D17108, doi:10.1029/2008JD010035.

Trigo, I.F., Boussetta, S., Viterbo, P., Balsamo, G., Beljaars, A., Sandu, I., 2015. Comparison of model land skin temperature with remotely sensed estimates and assessment of surface-atmosphere coupling. J. Geophys. Res. Atmos. 120, 12096Â–12111. https://doi.org/10.1002/2015JD023812.

Wan, Z., and J. Dozier (1996), A generalized split-window algorithm for retrieving land surface temperature from space, IEEE Trans. Geosci. Remote Sens., 34, 892–905.

Wan, Z., Y. Zhang, Q. Zhang, and Z.-L. Li (2002), Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sens. Environ., 83, 163-180.

Wan, Z. (2014) New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sens. Environ., 140, 36–45.

Wang, A., Barlage, M., Zeng, X., Draper, C.S., 2014. Comparison of land skin temperature from a land model, remote sensing, and in situ measurement. J. Geophys. Res. Atmos. 119, 3093Â–3106. https://doi.org/10.1002/2013JD021026.

Zheng, W., Wei, H., Wang, Z., Zeng, X., Meng, J., Ek, M., Mitchell, K., Derber, J., 2012. Improvement of daytime land surface skin temperature over arid regions in the NCEP GFS model and its impact on satellite data assimilation. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2011JD015901.