

Author Comment to Referee #1

We would like to thank the referees for their insightful and thorough review of the paper. All the referee's comments are addressed in the following response. The manuscript is being revised to accommodate these changes.

1) In this study, the microwave estimates of the LST diurnal cycle strongly rely upon the IR estimates. One expects the LST diurnal cycle to be impacted by clouds, even under partial cloud coverage along the day. The shape of the LST diurnal cycle (Holmes et al., 2015) will be affected if clouds are present at some time during the day. Can you clarify how these potential cloud effects are taken into account in the modeling / parameterization of the diurnal cycle? Two products are used in this study, the continuous MW-LST product and the MW-LST sparse product. It would be very interesting to see also results without any adjustments for the diurnal cycle. Would that be possible?

Reply: The influence of IR on the MW-LST estimates is solely through a time-constant linear scaling of the diurnal minimum, amplitude and timing. Therefore, this approach should not impact the presented results pertaining to the cloud effect on MW LST. But indeed, the diurnal cycle imposed on the MW-LST reflects an idealized clear sky situation. The result is that especially during partly cloudy days the fitted diurnal may not reflect the actual LST at the ground. Examples of this are apparent in Fig. 1 of the manuscript (eg June 24, 25, 28). However, the diurnal amplitude is determined based on a least-squares fit between diurnal model and daytime brightness temperature observations. This means that the overall diurnal increase in temperature is based on the observations, but not the hour-to-hour shape. As to the second part of the question; the difference between sparse observations and the diurnal temperature model is preserved in MW-LST-Sparse. In Fig 1. the effect of this is clear - MW-LST-Sparse corresponds more closely to the ground observations of LST than the continuous MW-LST. It is therefore used in this paper to present results with minimal effect of the diurnal model.

2) The number of MW-LST estimate is rather low as compared to what could be expected (50% with respect to 36% for the IR). Can you confirm that this is related to the need to have a very limited cloud coverage during the day, to anchor the microwave retrieval to the IR estimates? Would it be possible to relax this constraint, with modified assumption about the LST diurnal cycle (Holmes et al., 2015) under cloudy sky conditions?

Reply: The first column of Table 2 lists the overall data availability. The 50 % coverage for MW is indeed lower than what one would expect but this is not only due to cloud constraints. First of all the number is lower due the half year interval between AMSR-E and AMSR2 in 2012 (column 2 focuses on 2009-2011 for this reason). Secondly, the presented MW-LST excludes temperatures below freezing, and days where the temperature dips below freezing are excluded from the analysis. So column 4, with 64 % coverage of MW-LST in frost free days is closer to the general expectation of this method. The primary limitation in these conditions is the requirement of an observation near the daily maximum to constrain uncertainty of the diurnal amplitude. This effectively limits the method in its current implementation to days with an AMSR-E or AMSR2 ascending overpass: 2 out of 3 days. Incorporation of FenYun 3b could bring the coverage closer to daily.

3) The statistical metrics need to be better defined. The mean, the standard deviation, and the root mean square error are common statistical parameters that do not need to be explained. However, the ubRSME or the SEE need to be clarified. For instance, it was first understood that the SEE was the standard deviation, but from Table 4, it is obvious it is not the case (otherwise we would have $RMSE^2=STD^2+BIAS^2$). The advice is to use the standard and commonly accepted statistical metrics (mean, std, rmse), to avoid any ambiguity, or define clearly the metrics used.

Reply: We added a short section on Statistical metrics (2.4) that defines bias, RMSE, ubRMSE and SEE.

4) Clouds do affect the 37 GHz (p.10 | 31 sec. 4). The sentence 'Therefore Ka-band emission process' has to be removed or modified. Cloud particles do not scatter the radiation at 37 GHz (because of the small size of the particle with respect to the 37 GHz wavelength) but they do absorb and emit radiation.

This is actually how the cloud water content is estimated over ocean from observations at 37 GHz. Over land, cloud water retrieval is nevertheless very difficult, because of the lack of contrast between the cloud and the surface contribution to the signal (with an emissivity of the land surface close to 1 compared to the low emissivity of the ocean that provide a good contrast with the cloud emission).

Reply: Thanks for this insightful comment on the theoretical bases of the pathway through which clouds would affect Ka-band emission. We changed the sentence to: "Therefore, clouds do not scatter the 37 GHz radiation coming from the surface. They do however absorb and emit radiation themselves. The effect of clouds on MW-LST would thus be moderated by associated meteorological conditions like atmospheric vapor content and temperature profiles."

Minor points:

- P.2 I.27. 'Well validated TIR-LST'. Not that sure: Our recent experiences show significant differences between TIR-LST products that are all supposed to be well validated:

Reply. Point well taken. In the revised paper we focus the sentence on the geostationary aspect of the TIR-LST rather than its level of validation.

- P.2. I.27-28. Could you provide more information about the sources of these biases?

Reply: This scaling was able to account for biases in characteristics of the diurnal temperature cycle (DTC) related to Ka-band emissivity, sensing depth, and atmospheric effects (Holmes et al. 2015).

- P.3. I.23-24. How are the IR pixels handle within each 0.25 pixels? There are approximately 64 3km pixels in each 0.25 pixel. Do they all need to be clear to aggregate the data?

Reply. If two-thirds of the 3-km observations are masked than the sample average is rejected for that location and time.

- P.4. I.5-6. Can you specify here if the inter-calibration is in TB or LST space. The information might be available in other papers, but it needs to be reminded here.

Reply. We clarified in the text that the intercalibration is based on the brightness temperatures.

- P.5. eq. 4. It is not obvious here to realize how different is the MW-LST-Sparse from the 'raw' microwave LST. Can you comment on this point? Examples of the three MW-LST (continuous, sparse and 'raw') would help.

Reply. We added the following explanation below Eq. 4. "In MW-LST-Sparse the impact of the DTC3 model is limited to providing the minimum and amplitude of the diurnal. The differences between the observations and the diurnal model at the actual observation time are preserved. The difference between the continuous MW-LST and the MW_LST-Sparse is illustrated in Fig. 1 (lower panel). "

- P.7. I.10-11. Neighbouring daytime information taken for nighttime hours. That looks very uncertain. Information every 6 hours or every 3 hours as specified I9, but not used in I12-13? Did you try the analysis with only the daytime data?

Reply. We did do the analysis with only day-time data but this makes the results more influenced by biases in the diurnal amplitude. By using this estimate of cloudiness for night-time hours allows to use the night-time data.

- P.7. I.29. The 55% is after removal of the subfreezing temperatures? It will be good to give also the percentage only related to cloudiness.

Reply. The final two columns in Table 2 list the coverage for clear and cloudy sky. From this it can be seen that for TIR, the cloudy sky coverage is 15 % of clear sky coverage. For MW the cloudy sky coverage is 75% of the clear sky value. The 25 % gap between cloudy and clear sky coverage for MW is related to the spatial standard deviation filter applied to AMSR-E, TMI, and Windsat and this is thought to capture effects of active precipitation. Furthermore days where the difference between the sparse observations and the diurnal model are too large are removed.

- P.8 I.20. ThAn, not then

Reply. It has been changed.

- P.9. I.1-2. *Not that clear. MSG stations B,D,C are close to 1:1, as GOES stations J, H, M, G, E. Only K and J (GOES) are worse than D (MSG).*

Reply. Indeed. This explanation of increased scatter pertains to the ubRMSE.

P.9. I.29-30. How do the scaling of the MW-LST to the TIR-LST play a role here? It would be nice to see how the 'raw' MW-LSTs behave?

Reply. Matching the diurnal amplitude has a big effect on the efficacy of this method to detect cloud effects. If there is big mismatch between diurnal amplitude of site and satellite, then any variation in diurnal sampling between cloud bins will result in a bias that is superimposed on the cloud response signal.

With a longer data record such spurious effects may be mitigated but for now we think it's best to focus on the data sets in 'LST' space.

- P.11. Discussion. *It would be nice to see the linear regression of Table 4 also discussed in Figure 3, to see if the MW-LST cloudiness dependence (even small) changes by not having the MW-LS somehow depending on the IR diurnal cycle.*

Reply. Just as in our response to the comment above, the large influence of the bias in diurnal cycle with respect to the ground data makes it very complicated to analyze datasets with different diurnal biases. The 2009 regression does not account for the diurnal amplitude bias and will therefore underestimate the diurnal cycle in LST. It would be hard to disentangle the effect of changing sampling biases from actual cloud induces biases.

- P.11 I.9. *What is the spatial heterogeneity of the land surfaces for these two locations? That could be assessed with the std of the IR LST within the large 0.25 pixel.*

Reply. The spatial standard deviation of IR LST could indeed be a good indicator for spatial heterogeneity during clear sky periods, if one could make sure to avoid cloud artefacts in the data. However, it is not possible to get the same information for cloud covered scenes.

- Table1. *Something missing for station L. US-? Wha is the meaning of column W?*

Reply. The table is amended to clarify these meanings.

- Table 2. *CloudY not clouded.*

Reply. It has been changed.

- Fig1. *When the situation is almost clear (i.e. on July 26 on the upper panel), several IR values were expected, with a good description of the diurnal cycle. This is not the case (lower panel). Can you explain?*

Reply. We don't know for sure the reason behind these data gaps. As with any cloud mask, there could be false negatives as well as false positives. It may also be that significant cloud cover affected the 0.25 degree grid box but passed the station by.

- Fig1. *For the A calculation, in the top figure, it looks like a full diurnal cycle of SW is available. In the text, only 6 hour information and/or 3 hour information are mentioned.*

Reply. Shortwave radiation is reported at an hourly or half-hourly interval. The measure of cloudiness is calculated based on 3-hr intervals in order to provide a more time-averaged metric that might be more representative of the 0.25 degree spatial average.

- Fig1. *What is the meaning of the dashed black lines on July 29?*

Reply. The dashed line represents a case where the MW-LST was calculated but rejected based on a large difference between sparse observations and the diurnal fit. This wasn't explained in the text and we added it to the caption.

- Fig2. *Avoid the use of the ubRMSE or define it.*

Reply: We defined the use of ubRMSE more clearly in Section 2.4.

- Fig2. Can you comment on the differences between stations in terms of their soil moisture (or a proxy of the soil moisture if not available)? The amplitude of the diurnal cycle could actually be used.

Reply. This is an interesting observation. Locations D, I, J, K, L all have dry conditions with low vegetation. In Fig 2a all but (L) have a 0.5 to 1 k higher ubRMS for MW than for TIR. When there is less vegetation, the MW transmissivity is higher and the influence of soil emissivity on the observed Ka-band brightness temperature becomes larger. Small changes in soil moisture can affect the soil emissivity and will result in biases for the MW-LST when a constant emissivity is assumed (as in current implementation). This points to possible improvements when the scaling to TIR is performed at shorter window lengths, perhaps in 3-month moving windows. We added this discussion in the description of the Fig 2a in Section 3.1.

- Fig3. The indication of the cloud conditions with the circles is not straightforward. It would be helpful to add more explanation in the caption.

Reply. We added explanation in the caption: "For each site and cloud interval the percentage coverage of the temporal record is depicted in the top row with half-rounds in proportion to the number of data pairs. The potential number of data pairs (grey) refers to the number of in situ data points for each cloud bin. The actual number of data pairs is superimposed on this for MW (blue) and TIR (red)."

- Fig3. The spatial standard deviation of the IR LST information at each station might help explain some of the differences between stations, in terms of surface heterogeneity.

Reply. This would indeed be good additional information to assess the spatial complexity of LST for each station and might be something to take into account in future studies.

Author Comment to anonymous Referee #2

We would like to thank the referee for their thorough review of the paper. All the referee's comments (included below in italics) are addressed in the following response. The manuscript is being revised to accommodate these changes.

Reply to General comments:

1. Since there are other recent published studies on estimating LST with microwave observations such as André et al., (2015), Prigent et al., (2016), I suggest the authors integrate these studies in the introduction section.

Reply. The recent papers of André et al., (2015), Prigent et al., (2016) are now referenced in the introduction.

2. The study tested the hypothesis that microwave based LST is not sensitive to clouds over the FLUXNET tower scale. From the manuscript, it seems that the global scale microwave LST is already available. I suggest the authors to conduct a further study in the future on global analysis such as inter-comparison with other available LST products. Nevertheless, it would be nice if the authors could show the spatial pattern of the microwave LST and thermal LST. In addition, the current study focuses on the time period 2009-2011. For hydrological applications, this period is very limited. Does the developed microwave LST cover longer period? If not, do you have plans to extend it to long time period? I did not find detailed description on the global microwave product.

Reply. This study is indeed focused on testing the cloud tolerance of MW-LST by using detailed all-weather records of ground observations. At larger scales a detailed analysis of the diurnal characteristics and an assessment of random error over Europe, Africa and the middle-East was described in Holmes et al. (2015). In terms of global comparisons with other LST products and for longer timescales, we intend to make the dataset available to facilitate these studies in collaboration with other investigators.

3. The diurnal temperature cycle (DTC) of microwave LST is scaled to match that of thermal LST. Therefore, the diurnal microwave LST depends on the thermal LST to some extent. I am wondering how much influence would this scaling bring to the diurnal microwave LST. In other words, I suggest the authors show the results before scaling the DTC to match thermal LST.

Reply. The scaling of MW TB to TIR LST is done at per pixel basis and uses three parameters that are held constant over time: difference in diurnal timing, amplitude and daily minimum. These constant scaling parameters cannot affect the potential response to time-variant impact of clouds. If there are effects on the analysis than that would be through physical thresholds imposed (freezing point), or diurnal biases with the ground data that are sampled different with different cloud cover bins. This type of effect would add noise to our analysis but not fundamentally change the conclusions. Moreover, the fact that MW and TIR have such a different response to clouds is testimony to the independence of the MW-LST.

Reply to Specific comments:

1. Page 2 Line2: specify the name of the radiometer that deliver 2 km spatial resolution.

Reply: The Spinning Enhanced Visible and Infrared Imager (SEVIRI) is now given as an example with 3 km resolution.

2. Page 4 Line3: List the names of the 6 satellites/sensors.

Reply: These satellites include the Advanced Microwave Scanning Radiometer on EOS (AMSR-E) to October 2011 and its follow on AMSR2 from July 2012. Several platforms of the Special Sensor Microwave and Imager (SSM/I), the Tropical Rainfall Measurement Mission (TRMM) Microwave Imager (TMI), and Coriolus-WindSat.

3. Page 5 Line7: It would be nice if the used FLUXNET towers are shown in a global map.

Reply: We include two maps of the Fluxnet tower locations here (Figs 1 and 2, below), but we feel it is not of sufficient added value to the information in Table 1 to add to the manuscript.

4. The statistic metric R (correlation coefficient) should also be calculated except for $ubRMS$ and $BIAS$.

Reply. The Pearson correlation coefficient is not as discerning in the case of temperature when there is a strong seasonality that dominates the correlation. For the locations in this study the R values are between 0.94 and 0.99. MW and TIR perform very similar with only a few stations with a 0.01 lower R for MW. Only station J has a markedly lower R for MW-LST (0.91) Vs TIR (0.95).

Work cited:

Holmes, Thomas R. H., Wade T. Crow, Christopher R. Hain, Martha Anderson, and William P. Kustas. 2015. "Diurnal Temperature Cycle as Observed by Thermal Infrared and Microwave Radiometers." *Remote Sensing of Environment* 158C: 110–25. doi:10.1016/j.rse.2014.10.031.

Figures:



Fig 1. Map of European locations of fluxnet towers used in this study.



Fig 2. Map of US locations of fluxnet towers used in this study.

Summary of relevant changes made in the manuscript:

- Updated introduction with recent publications.
- Added a short section (2.4) on statistical metrics relevant to the paper.
- Added some discussion in (3.1) on the role of soil moisture to explain difference in unbiased RMSE between the sites, in answer to an insight-full suggestion by the reviewer.
- Changed the statement about cloud radiative properties (Section 4) according to reviewers comment

Cloud Tolerance of Remote Sensing Technologies to Measure Land Surface Temperature

Thomas R.H. Holmes^{1,2}, Christopher Hain³, Martha C. Anderson¹, Wade T. Crow¹

¹Hydrology and Remote Sensing Lab., USDA-ARS, Beltsville, MD, USA

5 ²Science Systems and Applications, Greenbelt, MD, USA

³Earth Science Interdisciplinary Center, University of Maryland, College Park, MD, USA

Correspondence to: Thomas R.H. Holmes (thomas.holmes@ars.usda.gov)

Abstract. Conventional means to estimate land surface temperature (LST) from space relies on the thermal infrared (TIR) spectral window and is limited to cloud-free scenes. To also provide LST estimates during periods with clouds, a new method was developed to estimate LST based on passive microwave (MW) observations. The MW-LST product is informed by 6 polar orbiting satellites to create a global record with up to 8 observations per day for each 0.25° resolution grid box. For days with sufficient observations, a continuous diurnal temperature cycle (DTC) was fitted. The main characteristics of the DTC were scaled to match that of a geostationary based TIR-LST product.

This paper tests the cloud tolerance of the MW-LST product. In particular, we demonstrate its stable performance with respect to flux tower observation sites (4 in Europe and 9 in the United States), over a range of cloudiness conditions up to heavily overcast skies. The results show that TIR-based LST has slightly better performance than MW-LST for clear sky observations but suffers an increasing negative bias as cloud cover increases. This negative bias is caused by incomplete masking of cloud covered areas within the TIR scene that affects many applications of TIR-LST. In contrast, for MW-LST we find no direct impact of clouds on its accuracy and bias. MW-LST can therefore be used to improve TIR cloud screening. Moreover, the ability to provide LST estimates for cloud-covered surfaces can help expand current clear-sky-only satellite retrieval products to all-weather applications.

1 Introduction

Information about the temperature of the land surface (LST) is an important element in the retrieval of many hydrological states and fluxes from satellite measured radiances. For example, the retrieval of soil moisture or precipitation from passive microwave observations requires a coincident estimate of LST (e.g. Owe et al., 2008). In other applications, the rate of change in temperature is contrasted with net radiation to estimate evaporation as a residual of the surface energy balance (e.g. Anderson et al., 2011).

The most direct way to estimate LST from space-borne instruments is by radiometers which measure within the thermal band of the infrared spectrum (TIR). Thermal emission within this frequency band can be related directly to the physical temperature of the land surface, and is more precisely termed the ensemble radiometric temperature (Norman and Becker,

1995). Space-borne TIR radiometers allow for very high spatial resolution imagery. Even at the height of geostationary platforms, ~~some~~ radiometers can deliver 32-km spatial resolution, eg the Spinning Enhanced Visible and Infrared Imager (SEVIRI). A drawback to the TIR technique is that - at such wavelengths - clouds completely block the emission from the land surface. This means that space-borne TIR radiometers give no information about the land surface below the clouds, and instead reflect the temperature and emissivity of the clouds. The result is that the quality of the cloud screening directly affects the quality of a TIR LST product.

An alternative, more cloud-tolerant technique is based instead on passive microwave (MW) observations. In particular the Ka-band (~37 GHz) was shown to have a strong link with LST (Owe and Van de Griend, 2001). Based on these observations, linear regression-based LST estimates were derived for the Ka-band (Holmes et al., 2009) and variants of these linear relations are currently used in soil moisture retrieval (Jackson et al., 1999; Owe et al., 2008). However, using a single linear regression across the globe ignores potentially significant differences in microwave emissivity and can result in large biases, especially in desert areas. It also cannot account for large differences in the amplitude of the diurnal cycle between MW- and TIR-based LST, which have been implicated in reduced soil moisture retrieval skill during daylight hours (Lei et al., 2015; Parinussa et al., 2011). In contrast to these relatively simple linear methods, a neural network method was developed by Aires et al. (2001) to estimate LST based on multiple microwave channels besides Ka-band. By using atmospheric and surface information in addition to TIR LST in the training of the scheme, this method minimized systematic bias in monthly mean temperatures. However, because the training is based on a single polar orbiting satellite, it cannot give diurnal temperature information. Both of these methods were compared with station data by Catherinot et al. (2011), giving strong confirmation on the lack of sensitivity of microwave-based LST estimates to cloud liquid water path. This method has recently been developed further to allow for more continuous application to a more diverse suite of satellites and overpass times possible (Prigent et al., 2016). One drawback to all passive microwave-based methods is relatively coarse spatial resolution as compared to TIR sensors. At Ka-band, the smallest footprint size currently achieved with polar orbiting satellites is 10 x 15 km.

Because of the complementary nature of TIR and MW-based LST, there is a clear interest in merging these two independent technologies. For example, TIR-LST based evaporation retrievals would benefit from observational data during cloudy periods (e.g. Anderson et al., 2011). On the other hand, microwave soil moisture retrievals from SMAP have the goal of 9 km spatial resolution and this poses a resolution challenge to MW-LST inputs if TIR-LST cannot be leveraged. Reconciling the systematic differences in diurnal temperature cycle (DTC) between TIR and Ka-band is ~~As a~~ first step towards an ultimate merger of diurnally continuous TIR and MW LST products, To do this, Holmes et al. (2015) developed a method to scale the diurnal characteristics of multi-satellite dataset of Ka-band observations to TIR-LST. ~~B to~~ because a well-validated TIR-LST products exist with 15-minute temporal sampling from geostationary satellites, ~~T,~~ this scaling was able to account for biases in characteristics of the ~~diurnal temperature cycle (DTC) related to Ka-band emissivity, sensing depth, and atmospheric effects (Holmes et al., 2015)~~. By explicitly taking account of systematic differences in DTC between TIR and Ka-band, this method is able to estimate LST at any time of day from sparse Ka-band observations. Note that a similar pixel-

[by-pixel approach was applied by André et al \(2015\) over the arctic region where with a single satellite can provides diurnal sampling.](#)

The aim of this paper is to evaluate the new global MW-LST dataset in comparison to existing TIR-LST data over clear-sky days and [particularly](#) to test the assumption that MW-LST is tolerant to high levels of cloud coverage. Ground observations provide a common benchmark to test the relative accuracy of the two satellite products. Because the diurnal MW-LST product (Holmes et al., 2015) was scaled to TIR-based LST as produced by the Land Surface Analysis Satellite Application Facility (LSA-SAF, see <http://landsaf.meteo.pt>), the evaluation is mostly concerned with temporal precision, not with absolute bias. In previous work it was shown that relative aspects of a coarse-scale product can be evaluated using sparse in situ observations (Holmes et al., 2012). For a thorough discussion of absolute accuracy, readers are referred to papers detailing validation exercises for LSA-SAF-LST ([e.g. Ermida et al., 2014; Göttsche et al., 2013, 2016](#)) ([Ermida et al., 2014; eg. Göttsche et al., 2013](#)).

After establishing the accuracy of MW-LST relative to TIR-LST for a particular site, the stability of the precision of MW-LST (relative to ground data) for increasing cloudiness will be tested. Previous work showed indications of cloud tolerance of MW-LST in comparison to TIR-LST (Holmes et al., 2015), but the analysis used proxies for both cloud cover and LST quality. In this paper we use a more direct estimate of cloudiness and provide a more detailed look at the validation statistics for different levels of daytime cloudiness with the ground station as the reference. The hypothesis we test is that clouds affect a satellite measured LST by introducing an error (E). If E is consistent in sign throughout the measurement period (e.g. if clouds always lower the satellite LST estimate) this will introduce a systematic bias that will increase with cloud cover. If on the other hand the sign of E varies it will increase the random error in LST, but not necessarily a systematic bias. Only if we do not see a systematic bias with increasing cloud cover, nor an increase in random error can we reject the hypothesis that clouds affect the satellite LST.

2. Materials

2.1 Satellite LST estimates: Thermal infrared

TIR-LST is available from many sources, including both low-earth orbiting satellites and geostationary satellites. Because of our interest in the diurnal features of LST, used in surface energy balance evaluations, this study focuses on TIR-LST products developed from geostationary satellites. The first product is based on the ~~Spinning Enhanced Visible and Infrared Imager (SEVIRI)~~ [SEVIRI](#) onboard the Meteosat Second Generation (MSG-9) satellite. MSG-9 is positioned over the equator at 0° longitude. It has geographic coverage of Africa, Europe and the east coast of South America (with incidence angles below 70°). The Land Surface Analysis, Satellite Application Facility (LSA SAF) produces operational LST products based on split window observations (channels centered at 10.8 μm and 12.0 μm) of MSG-9. LSA SAF LST is originally produced

at 3 km spatial resolution. For this study, the data is aggregated to match the 0.25° resolution of MW-LST. If two-thirds of the 3-km observations are masked than the sample average is rejected for that location and time.

For North-America, NOAA operates the Geostationary Operational Environmental Satellites (GOES). GOES Surface and Insulation Products (GSIP) are produced by the office of Satellite Data Processing and Distribution, NESDIS, NOAA, and includes LST (V3 was used for this study). Unlike MSG-based LST, GOES GSIP LST is based on a dual window technique (3.9 and 11.0 μm), rather than the preferred split window technique due to the lack of a 12 micron thermal channel on the current generation of the GOES imager. An operational hourly LST at 0.125° degree spatial resolution is available from 2 April 2009 onwards. For this study, we averaged the nine 0.125° degree nodes that cover the edge and center of the 0.25° MW-LST product grid cell. In order to further reduce possible cloud contamination, a particular data point is only used if all nine 0.125° degree pixels covering the 0.25° grid box contain (non-cloud-masked) data.

2.2 Satellite LST estimates: Microwave

The MW-LST product is based on vertical polarized Ka-band (36-37 GHz) brightness temperature (T^{Ka}) as measured by microwave radiometers on 6 satellites in low earth orbit. These satellites include the Advanced Microwave Scanning Radiometer on EOS (AMSR-E) to October 2011 and its follow on AMSR2 from July 2012. Several platforms of the Special Sensor Microwave and Imager (SSM/I), the Tropical Rainfall Measurement Mission (TRMM) Microwave Imager (TMI), and Coriolus-WindSat. These observations are combined to create a global record with up to 8 observations per day for each 0.25° resolution grid box. The data are binned in 15 minute windows of local solar time (0:00-0:15 is first window of the day). The brightness temperatures are, and inter-calibrated using observations from the TRMM satellite (with an equatorial overpass) as a transfer reference. Individual 0.25° averages are masked if the spatial standard deviation of the oversampled Ka-band observations exceeds a prior determined threshold for a given grid box. Both the inter-calibration and quality control procedures are described in detail in Holmes et al (2013a).

The methodology to estimate LST from this record of Ka-band observations is described in (Holmes et al., 2015) and summarized below. For days with suitable observations (a minimum of 4, including at least one within a third day length from solar noon) and no $T^{Ka} < 250 K$ (an indication of frozen soil), a continuous diurnal temperature cycle (DTC) is fitted. The DTC model used is based on Göttsche and Olesen (2001) with slight adaptations to limit the number of parameters. This implementation (DTC3) is fully described in Holmes et al (2015). DTC3 summarizes the DTC with two daily parameters (daily minimum T_0 at start and end of day, and diurnal amplitude A) together with diurnal timing (φ), that is assumed a temporal constant (Holmes et al., 2013b). The daily mean is defined by the daily minimum and the amplitude ($\bar{T} = T_0 + A/2$). The Ka-band DTC parameters for individual days (\bar{T}_d^{Ka}, A_d^{Ka}) are scaled to match the long term mean of TIR observations:

$$A_d^{MW} = A_d^{Ka} / \delta \quad (1)$$

$$\bar{T}_d^{MW} = \beta_0 + \beta_1 \bar{T}_d^{Ka} \quad (2)$$

The scaled parameters are indicated with the superscript ‘MW’. The parameter δ represents the slope of the zero-order least squares regression line for estimating the amplitude of TIR-LST (A_d^{TIR}) from A_d^{Ka} . The intercept (β_0) and slope (β_1) to correct the mean daily temperature (\bar{T}_d^{Ka}) for systematic differences with TIR-LST (\bar{T}_d^{TIR}) are determined with a constrained numerical solver, as in Holmes et al. (2015). The constraint is based on radiative transfer considerations and assures that the scaling of the mean is in agreement with the prior scaling of the amplitude (Eq. 1). These scaling parameters were determined for each 0.25° grid box based on data for the period 2009-2012. The scaling (Eqs. 1 and 2) is applied to every day for which estimates of \bar{T}_d^{Ka} and A_d^{Ka} are available. Together with the timing of the diurnal cycle of TIR-LST, φ^{TIR} , as determined based on (Holmes et al., 2013b), we then calculate the diurnal MW-LST based on the same DTC3 model:

$$10 \quad \text{MW-LST} = \text{DTC3}(\varphi^{TIR}, T_0^{MW}, A^{MW}) \quad (3)$$

Comparing actual Ka-band observations to estimates provided by the fitted DTC model provides a valuable means of quality control. The root mean square error (RMSE) of the misfit between the DTC3 model and the sparse T^{Ka} observations is used to flag days where the assumptions imposed by the shape of clear sky DTC are not valid or individual Ka-band observations have a large bias.

Besides the continuous MW-LST product, we can also evaluate the product at the actual Ka-band observation times (thus weakening our reliance on the DTC3 model). To do this we project the difference between the original MW data and the DTC model fit onto MW-LST. -This product is referred to as MW-LST-Sparse:

$$20 \quad \text{MW-LST-Sparse} = \text{DTC3}(\varphi^{TIR}, T_0^{MW}, A^{MW}) + \frac{\langle A^{TIR} \rangle}{\langle A^{Ka} \rangle} (T^{Ka} - \text{DTC3}(\varphi^{Ka}, T_0^{Ka}, A^{Ka})) \quad (4)$$

In MW-LST-Sparse the impact of the DTC3 model is limited to providing the minimum and amplitude of the diurnal. The differences between the observations and the diurnal model at the actual observation time are preserved. The difference between the continuous MW-LST and the MW LST-Sparse is illustrated in Fig. 1 (lower panel).

2.3 Ground observations

25 FLUXNET is a worldwide network of meteorological measurement towers (flux tower) with common measurement protocols (Baldocchi et al., 2001). Each flux tower includes an instrument positioned above the vegetation canopy to measure net radiation. This instrument is made up of two pyranometers to measure up and down welling short wave radiation and two pyrgeometers to measure up and down welling long wave radiation. The radiometric surface temperature (T) can be derived from the long wave radiation measurements (upwelling: L^\uparrow , Wm^{-2} , and downwelling: L^\downarrow , Wm^{-2}) using the following equation:

$$30 \quad L^\uparrow = \varepsilon_L \sigma T^4 + (1 - \varepsilon_L) L^\downarrow \quad (5)$$

where ε_L is the broadband emissivity over the spectral range of the pyrgeometer (4.5-42 μm), σ is the Stefan-Boltzmann constant ($\sigma=5.670 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-1}$).

2.3.1 Long wave emissivity

5 An estimate of ε_L is obtained for each site based on measurements of L^\uparrow together with additional measurements of screen level air temperature (T_a), sensible heat flux (H) and wind speed (u). This estimate is based on three assumptions: 1) H is directly proportional to the near surface temperature gradient, 2) the difference $T - T_a$ represents that temperature gradient, and 3) $H = 0$ when there is no vertical temperature gradient (i.e. $T = T_a$). With these three assumptions in mind we then iterate over ε_L to find the solution where the regression of H against $(T - T_a)$ goes through zero and the squared errors are

10 minimized. When measurements of wind speed are available they are used to select atmospheric conditions where the relationship between H and the near surface temperature gradient is strongest ($u > 2 \text{ m/s}$). For forest sites, the direct relationship between H and T gradients breaks down. In those cases the simpler assumption is used that the long term average of T and T_a are equal: $\langle T \rangle = \langle T_a \rangle$. For more discussion and examples of this method see Holmes et al. (2009). In this study we apply this method to determine monthly ε_L for each site individually, and then use the median value of ε_L

15 (listed in Table 1) to calculate T based on Eq. 5. The standard deviation of the monthly measurements of ε_L are also listed in Table 1 and provides an indication of both uncertainty and seasonal variation in ε_L .

2.3.2 Spatial representativeness and selection tower sites

The tower-based estimate of T , from (Eq. 5), directly represents only the immediate tower surroundings within a radius of approximately 50 m. Clearly this is a very small spatial sampling of the 0.25° grid box ($\sim 25 \times 25 \text{ km}$) represented by the

20 satellite LST estimates used in this study (see Section 2.1 and 2.2). As a consequence, we typically find large systemic differences between the station data and the areal average. Given that overall weather conditions are relatively homogenous over distances of 25 km, these differences can be attributed to the land cover type at the station location in comparison to that over the entire grid box (for examples of this, see Holmes et al. (2009)). The representation of the spatial average by ground observations can be improved significantly if more than one station is available in the same grid box and the towers

25 are situated in thermodynamically contrasting land cover types: (forest and cropland/herbaceous). In that case the land cover associated with the tower site(subscript, s) determines the weight (W) according to the spatial fraction of that land cover type within the 0.25° grid box (MCD12C1, Friedl et al. (2010)). We use this information to estimate the grid average LST as the weighted average of site measured T according to Eq. (6):

$$LST = \frac{\sum_{s=1}^n W_s T_s}{\sum_{s=1}^n W_s} \quad (6)$$

For example, site DE-Hai of location A is located in a forest and represents 16 % of the pixel. Site DE-Geb is located in croplands and represents the 80 % of the pixel that has low vegetation cover or bare soil. Urban, open water, or wetland account for the remaining 4 %, which does not affect the weighting. We only considered locations where this rest fraction is below 5 % of the grid coverage. Another criterion for site selection was that the land cover at the site must represent more than 75 % of the pixel. Sites in mountainous areas are also excluded. For the period of 2009-2012 this means there are 13 locations with at least 2 years of flux tower sites available for this study, and four of these locations contain multiple stations. For locations where only one stations is available, LST is set equal to the site measurement: $LST = T$. All the validation targets are listed in Table 1, together with the geographic location of the individual stations, the land cover type as reported by the flux tower operators and the parameters W and ε_L as described above.

10 2.3.3 Cloudiness at tower location

The down welling short wave radiation (S^\downarrow) as measured at the fluxtower is strongly affected by the amount of condensed water in the atmosphere. We can therefore use the reduction in site measured daytime S^\downarrow relative to an expected value during clear skies as a proxy for cloudiness. The clear-sky irradiance S_{clear}^\downarrow is estimated based on top-of-atmosphere solar irradiance (S^{TOA}) which can be calculated based on geographic location and day of year (Van Wijk and Ubing, 1963). Even on a clear day, atmospheric absorption reduces the irradiance at the surface by 20 to 30 % from the top of atmosphere value. We estimate this clear sky absorption (A_{clr}) by calculating the slope (β) of the zero order linear regression between S^\downarrow and S^{TOA} for days that are in the highest quintile of S^\downarrow/S^{TOA} : $A_{clr}=1-\beta$. These estimates of A_{clr} (listed in Table 1 for each individual site) range from 0.22 to 0.31 and show a good agreement between stations of the same cluster. We use the minimum recorded value for each cluster to calculate S_{clear}^\downarrow :

$$20 \quad S_{clear}^\downarrow = S^{TOA}(1 - A_{clr}) . \quad (7)$$

By using S_{clear}^\downarrow to normalize measured S^\downarrow , we account for solar zenith effects and can formulate a measure for shortwave cloud absorption (A_{cloud}), expressed in percentage:

$$A_{cloud} = 100 \frac{S_{clear}^\downarrow - S^\downarrow}{S_{clear}^\downarrow} \quad (8)$$

This definition of A_{cloud} is used as a measure of cloudiness and calculated based on 3 hour totals of insolation for the daytime between 6AM to 6PM. Obviously this ~~approach-definition of cloud absorbtion~~ does not ~~work-apply for night-~~ ~~timewhen the sun is below the horizon~~. For night-time hours we use the ~~neighbouringneighboring~~ daytime window:

$$A_{cloud}(0 - 6) = A_{cloud}(6 - 12) \quad (9)$$

$$A_{cloud}(18 - 24) = A_{cloud}(12 - 18) \quad (10)$$

Figure 1 gives an example of the site measured S^\downarrow and the calculated A_{cloud} for an 8-day summer period at station B (top panel). The bottom panel shows the site measured LST and illustrates how the temporal sampling of the satellite products is affected by clouds.

2.4 Statistical metrics

In the description of the results we make use of standard statistical metrics. In terms of absolute error metrics we report bias, the long term mean difference between satellite product and in situ data, and the root mean square error (RMSE). By removing the long term mean difference we can calculate RMSE of the unbiased data (ubRMSE). These three metrics are related as follows:

$$RMSE = \sqrt{ubRMSE^2 + bias^2} \quad (11)$$

We further report standard error of estimate (SEE) as a measure of temporal precision:

$$SEE = \sigma \sqrt{1 - \rho^2} \quad (12)$$

where σ is the standard deviation of in situ data and ρ is Pearson's correlation coefficient.

3. Results

We acquired data for 17 field sites in 13 unique grid locations with data records within the 4-year time period of 2009 to 2012. Table 2 lists the amount of days with at least 12 hours of observations for either MW or TIR-LST to give an overall sense of available validation data for this study. In total we have 13316 data days of in situ data (36 out of 44 data years). Of these data days, 50 % also have MW-LST estimates, and 36 % have TIR-LST observations. For MW, this percentage is negatively affected by the gap between AMSR-E (radiometer turned off on October, 2011) and AMSR2 (first observations in July 2012). The MW-LST product is heavily-reliant on these satellites with a midday overpass for constraining the diurnal amplitude. For TIR, the percentage is particularly low for GOES due to a more stringent cloud filter than employed for MSG.

To better represent the relative data coverage that is possible with the two LST retrieval techniques, we focus on the four station pairs in the MSG domain and limit the time period to 2009-2011. We further focus on the days where the minimum temperature (as measured at the station) stays above freezing (the MW method is not applicable for subfreezing temperatures). Within this smaller subset, we have 2506 data days of in situ data and the coverage of MW is 55 % in comparison with the 42 % coverage for TIR. However, breaking this down by cloud cover reveals the big difference in coverage resulting from the wavelength-dependant tolerance to clouds. During clear skies, the coverage of TIR is 93 %, and MW comes in at 76 % (mainly attributed to the 2/3 days coverage for AMSR-E). During cloudy days the coverage drops to 13 % for TIR, whereas for MW it maintains 55 % coverage.

In the following section we want to answer two questions. How does the MW LST compare to TIR LST in relation to ground data during days with clear skies? And is the performance of MW-LST affected by clouds? We focus on hourly average temperatures for days where the station data remains above 1°C to avoid snow or frozen surface conditions.

3.1 Clear-sky comparison of satellite LST products

The ground observations provide a common benchmark to test the relative accuracy and bias of the two satellite products with the same in situ data. Days with clear skies are selected based on the measure of cloudiness as defined in Section 2.3.3, with a maximum accepted value of $A_{\text{cloud}} = 0.2$. This is in addition to the cloud screening performed in the generation of the TIR products (Section 2.1), the quality control of the MW LST (Section 2.2), and the selection of frost-free days. Even though the spatial representativeness and uncertainty in ε_L may insert systematic errors in the estimation of the spatial average from the ground stations, they can be used as a reference to compare different satellite LST products.

For each of the 13 validation targets we tabulate ~~(see Table 3) both unbiased RMSE (ubRMSE) and bias (see Section 2.4 for definition) (see Table 3).~~ By excluding long term bias. Most interesting is the ubRMSE, which gives an indication of the overall data quality, which. ~~This statistic~~ includes the random error and errors resulting from a mismatch in variance (either seasonal or diurnal), ~~but not the overall bias~~. Errors in spatial representation of the sites affect MW and TIR in the same way. In order to highlight the relative performance of the two satellite products with respect to the common benchmark, we compare their performance directly in Fig. 3.

Encouragingly, the multi-site average ubRMSE for the four European Fluxnet sites shows the MW-LST (2.3 K) to be only moderately higher ~~then than~~ TIR-LST (2.1 K) for these frost-free and cloud-free observations. This is a positive result for MW-LST because of the extra processing needed to correct MW data for sensing depth differences with TIR. Both satellite products have a higher ubRMSE with the Ameriflux stations, but again the multi-site average ubRMSE for MW-LST (3.0 K) is only slightly higher than that for TIR-LST (2.8 K). Figure 2a compares the ubRMSE with in situ data directly for the two satellite technologies. The high correlation between the two methods is an indication that spurious effect of spatial representation of the site affects both methods to similar degrees. Of all the stations, MW-LST has a lower ubRMSE at 5 of the 13 stations and the only stations where we record more than 0.5 K difference in ubRMSE between TIR and MW-LST are Fluxnet station D (2.5 K for MW Vs. 1.7 K for TIR) and Ameriflux station I (3.5 K for MW Vs. 2.9 K for TIR) and J (3.4 K for MW Vs. 2.4 K for TIR). These stations, together with K and L, all have dry conditions with low vegetation. When there is less vegetation, the influence of soil emissivity on the observed Ka-band brightness temperature becomes larger. Small changes in soil moisture can affect the soil emissivity and will result in biases/errors in ~~for~~ MW-LST when a constant emissivity is assumed (as in current implementation). This points to possible improvements when the scaling to TIR is performed at shorter window lengths, perhaps in 3-month moving windows.

Because MW-LST is scaled directly to TIR-LST its bias is almost completely determined by the bias between TIR-LST and the site (Table 3 and Figure 2b.). The European Fluxnet sites fall within the MSG domain, and these data years were part of the data on which the scaling of MW-LST is trained (Holmes et al. 2015). ~~The TIR LST in the domain of the Ameriflux sites is based on either GOES East or GOES West data, and is produced based on a 3.9/11 dual window technique rather than the more typical 11/12 micro split window available to the MSG-based LSA SAF LST. Together this accounts for increased~~

~~scatter in bias between the two satellite products and the ground data when comparing MSG domain (circles) and GOES domain (squares).~~

Although the mean bias (Fig 2b) is almost identical, the bias in morning heating (ΔT , Fig 2c) has more variation between the two satellite products. It is interesting that generally the satellite products overestimate ΔT compared to ground data: on average they both overestimate the recorded heating at the stations by about 10 % .

3.2 Cloud Tolerance of Satellite LST

To test the stability of the MW-LST for increasing levels of cloudiness we took a closer look at the 4 sites in Europe and 3 in the US (site A-G). To isolate the effect of clouds on the agreement between satellite and ground observations, we first remove structural differences by fitting a linear regression for each location, based on data with cloudiness below 20 % (Γ – see section 2c). We then divide the data in 5 equal bins of increasing cloud coverage from 0 to 100 %. The RMSE and mean difference (bias) between the satellite data and the regression-corrected in situ data is then calculated for each 20 % cloud bin. The purpose is to test the assumption that MW-LST is tolerant to higher levels of cloud coverage.

Figure 3 shows the result of this analysis for location A-G (from left to right). For each location the data coverage (top row), RMSE (middle row), and bias (bottom row) are displayed for the five 20% bins of cloudiness. First of all, the large increase in negative bias with increasing cloudiness for the TIR-LST product stands out. At all stations we see a clear negative bias in response to increasing cloudiness for TIR-LST, and the overall agreement between stations is striking. At 40-60% cloud cover, all but one station show a significant negative bias for TIR-LST. Above 60 % cloud cover all stations (where TIR-LST is still available, presumably due to failure of the cloud mask) show a negative bias of 2 K or more. This clearly shows that for TIR-LST we have to accept the hypothesis that clouds affect the satellite LST estimate, even after a cloud mask is applied. It is well known that TIR observations are sensitive to clouds and that a failure to mask for cloud conditions will result in an underestimated LST (for land surface above freezing). Because of this systematic response to clouds, the bias metric by itself is a good indicator of the effect of cloud contamination in clear sky TIR-LST products. The symbols in the top row show the diminishing temporal sampling with increasing cloud cover. When we contrast this with the size of the bias it is clear that the cloud mask as implemented in the LSA-SAF product (for sites A-D) is not sufficient at removing cloud artifacts. The GOES product (for sites E-G) appears to remove times with high cloud values more completely. Although investigating the efficacy of cloud masks for TIR techniques is not the purpose of this paper, it does help illustrate how cloud effects can be identified with these ground stations.

In clear contrast to the TIR-LST products, the response of MW-LST to increasing cloudiness is much more muted and not as consistent across stations. Stations A, B, C, and E show no response in terms of bias and below 80 % cloud cover there is no station with a MW-LST bias of more than 1 K. One station shows a negative trend (D), and two stations show a positive trend (F and G). But only above 80 % cloudiness do these trends result in bias error of greater than 1 K. Because we see both positive and negative biases in the MW LST analysis we cannot rely solely only on the bias metric to assess the impact

of clouds. If there are cancelling biases affecting an individual station, this could suppress the bias. The increased retrieval error would still be reflected in an increased RMSE. However, the RMSE of MW LST changes minimally relative to its baseline value at 0-20 % cloudiness, and mirrors the size of the bias. This indicates that there is little potential for ‘hidden’ biases behind these numbers. For MW-LST we can therefore reject the hypothesis that clouds affect the satellite LST estimate.

The MW-LST-Sparse product (Eq. 4) adopts the same scaling with TIR as the diurnal MW-LST but has much less sensitivity to the imposed shape of the diurnal model (DTC). For clear skies this distinction is negligible, as apparent from the almost identical values of ubRMSE shown in Table 3. The effect of the clear sky model is likely to be higher in days with cloudy or partial cloud covered sky. And although the sparse set only has 4-8 observations per day, it allows more samples in days with complex temperature changes. Such days are removed from the MW-DTC product if no good match is found between the diurnal model and observations. We can therefore use the MW-LST-Sparse product to test for undue influence of the DTC model (and its related quality flags) on the relationships between LST errors and cloudiness. The response in bias of MW-LST-Sparse to increasing cloudiness is almost identical to the response of MW-LST for each station (see Fig. 3). In terms of RMSE the sparse set shows equal or higher values than the diurnally-continuous MW-LST product, which is not surprising as it does not have the smoothing and quality control associated with the DTC model.

3.3 All-sky validation by satellite overpass time

The MW-LST record is a combination of different satellites. In the following analysis the validation results of the MW-LST product are broken down by time of day and satellite input record. All data pairs where the minimum temperature at the station stays above freezing are included in this analysis, regardless of cloud cover. It is interesting to compare these results to the much simpler approach that uses a single linear regression model globally (Holmes et al., 2009). Table 4 lists RMSE, ~~standard error of estimate (SEE, precision)~~ and bias for the old and new approach. The statistics are aggregated for all locations as listed in Table 1. The mean scaling with TIR-LST results in a drop in bias for the MW-LST, reducing the average RMSE by 1 K. Part of this reduced RMSE results from the improved characterization of the amplitude of the diurnal cycle, which improves the slope at all times of day and accounts for 0.2 K of the improvement in RMSE. The impact on the ~~random error component~~ precision (quantified here by SEE) is mixed – on average there is no change. Biggest improvements in all metrics are recorded for the forest locations (Sites C, E, and H).

4 Discussion

Considering all 8 locations used in the cloud analysis we see little to no response to clouds in terms of bias and RMSE for MW-LST and this allows us to reject the hypothesis that clouds negatively affect its accuracy. However, for three sites we do find weak and opposing biases at higher cloud coverage which require an explanation. The wavelength of Ka-band (8 mm) is two orders of magnitude larger than a typical cloud droplet (10 μm). [Therefore, clouds do not scatter the 37 GHz radiation](#)

~~coming from the surface. They do however absorb and emit radiation themselves. The effect of clouds on MW-LST would thus be moderated by associated meteorological conditions like atmospheric vapor content and temperature profiles. Therefore, any effect of clouds on MW-LST would stem from changes in associated meteorological conditions like atmospheric vapor content and temperature profiles and their potential impact on Ka-band emission processes.~~ According to

5 the zero-order radiative transfer model, an increased atmospheric opacity (through increasing atmospheric water content) increases the weight of the atmospheric contribution to the satellite measured brightness temperature, relative to the top of vegetation emission. The sign and size of the effect of a change in atmospheric opacity thus depends on the contrast between the atmospheric temperature and the land surface temperature times the effective emissivity. It is therefore possible that this could explain the site-to-site differences in bias as shown in Fig 3. Analyzing the overall effect of the atmosphere on biases
10 in MW-LST will require more detailed atmospheric profile information coupled with a radiative transfer model.

Another possible explanation is that the positive biases recorded at locations F and G are related to scale differences between the site and the 0.25 degree grid cell. Spatial heterogeneity in LST is likely more pronounced during clear-sky periods when spatially varying soil and vegetation ~~wield-yield~~ a strong influence on the day time temperature gradients. During cloudy
15 periods the temperature gradients are not as pronounced and more directly linked to the more uniform air temperature. If the mean temperature at the station is generally higher than the areal mean LST, and this bias diminishes with increasing cloudiness, this would be transformed through our clear-sky training into a positive bias for the satellite product at high cloudiness. However, this effect would affect both MW and TIR to the same extent. We have tested this at locations with two stations in contrasting land cover types (A-C). What we found is that indeed it is possible to ‘rotate’ the bias response by changing the weights of the individual stations, and that this rotation affects both MW and TIR-LST. This effect of site
20 representation can therefore explain the greater variation in response from station to station for locations where only one station was available (D-G).

5 Conclusion

In this paper, a recently developed satellite MW-LST product is compared to ground station data and satellite TIR-based LST products. The MW-LST was developed to complement TIR-LST with a coarser spatial resolution but at a higher
25 temporal resolution. The higher temporal resolution of MW-LST is based on the assumption that MW has a relatively high tolerance to clouds, which allows for observations at times when no TIR observations are possible. This paper tests this assumption by looking at the precision with respect to ground stations for increasing levels of estimated cloudiness. Our analysis is performed at the 0.25° spatial resolution as predicated by the MW-LST product. At this coarse spatial resolution, the overall unbiased RMSE between TIR-LST and ground stations during clear sky days is 2.1 K for the four locations in the
30 MSG domain, and 2.8 K for the 9 locations in the GOES domain. For the same locations we find that the MW-LST is only slightly higher (+0.2 K for both domains).

With increasing cloudiness the RMSE increases significantly for TIR-LST, caused by a matching negative trend in bias that is seen at all seven locations. This demonstrated the known effect that clouds have on TIR estimates of LST. The fact that these trends are so apparent highlights the limitations of current cloud screening techniques as employed in TIR-LST products that are in general use. In clear contrast to this we find a much more limited response in both RMSE and bias for
5 MW-LST. Because of this we conclude that there is no significant direct impact of clouds on the accuracy of the MW-LST product. However, at three stations the size and sign of the response is such that further research is needed to identify the exact causes introducing error in MW-LST. By taking into account the atmospheric humidity and temperature profile further analysis may investigate the extent to which this mixed response can be explained by atmospheric conditions associated with cloudiness. Alternatively, if a greater database were available of locations with flux tower sites in contrasting land covers
10 this could be used to isolate the role of scale mismatch between station and the satellite product.

As an immediate outcome the result of this work highlights the utility of MW technology for cloud screening of TIR-LST. This is something that will be explored in future work. Ultimately, the goal is to find the best way of combining MW and TIR technology for the estimation of LST from space.

Data Availability

15 Timeseries of MW-LST and TIR-LST covering the locations and time period of this paper are available upon request from the corresponding author. The global source data for MW-LST are publicly available. They are aggregated from several data centers and we would like to thank Goddard Earth Sciences (GES) Data and Information Services Center (DISC) for archiving and distributing TRMM satellite as acquired by NASA's Earth-Sun System Division, National Snow and Ice Data
20 Center for archiving and distributing Aqua-AMSR-E data, and NOAA's Comprehensive Large Array-data Stewardship System (CLASS) for dissemination of Defense Meteorological Satellite Program data. LSA SAF disseminates EUMETSAT products. This work further used data acquired by the FLUXNET community (fluxnet.ornl.gov) and in particular by the following networks: AmeriFlux and CarboEuropeIP.

Acknowledgements

This work was funded by NASA through the research grant "The Science of Terra and Aqua" (13-TERAQ13-0181). We
25 would further like to thank Li Fang (NOAA) for preparation and interpretation of GOES LST.

References

Aires, F., Prigent, C., Rossow, W. B. and Rothstein, M.: A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations, *J. Geophys. Res.-Atmospheres*, 106(D14), 14887–14907, doi:10.1029/2001JD900085, 2001.

- Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., Gonzalez-Dugo, M. P., Cammalleri, C., d'Urso, G. and Pimstein, A.: Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery, *Hydrol. Earth Syst. Sci.*, 15(1), 223, 2011.
- 5 André, C., Ottlé, C., Royer, A. and Maignan, F.: Land surface temperature retrieval over circumpolar Arctic using SSM/I–SSMIS and MODIS data, *Remote Sens. Environ.*, 162, 1–10, doi:10.1016/j.rse.2015.01.028, 2015.
- 10 Baldocchi, D., Falge, E., Gu, L. H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X. H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K. and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities., *Bull Am Meteorol Soc*, 82, 2415–2434, 2001.
- Catherinot, J., Prigent, C., Maurer, R., Papa, F., Jiménez, C., Aires, F. and Rossow, W. B.: Evaluation of “all weather” microwave-derived land surface temperatures with in situ CEOP measurements, *J. Geophys. Res. Atmospheres*, 116(D23), n/a–n/a, doi:10.1029/2011JD016439, 2011.
- 15 Ermida, S. L., Trigo, I. F., DaCamara, C. C., Göttsche, F. M., Olesen, F. S. and Hulley, G.: Validation of remotely sensed surface temperature over an oak woodland landscape—The problem of viewing and illumination geometries, *Remote Sens. Environ.*, 148, 16–27, 2014.
- Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote Sens. Environ.*, 114(1), 168–182, 2010.
- 20 Göttsche, F.-M. and Olesen, F. S.: Modelling of diurnal cycles of brightness temperature extracted from METEOSAT data, *Remote Sens. Environ.*, 76(3), 337–348, doi:10.1016/S0034-4257(00)00214-5, 2001.
- Göttsche, F.-M., Olesen, F.-S. and Bork-Unkelbach, A.: Validation of land surface temperature derived from MSG/SEVIRI with in situ measurements at Gobabeb, Namibia, *Int. J. Remote Sens.*, 34(9–10), 3069–3083, doi:10.1080/01431161.2012.716539, 2013.
- 25 Holmes, T. R. H., De Jeu, R. A. M., Owe, M. and Dolman, A. J.: Land surface temperature from Ka band (37 GHz) passive microwave observations, *J. Geophys. Res. Atmospheres*, 114, D04113, doi:10.1029/2008JD010257, 2009.
- Holmes, T. R. H., Jackson, T. J., Reichle, R. H. and Basara, J. B.: An assessment of surface soil temperature products from numerical weather prediction models using ground-based measurements, *Water Resour. Res.*, 48(2), W02531, doi:10.1029/2011WR010538, 2012.
- 30 Holmes, T. R. H., Crow, W. T., Yilmaz, M. T., Jackson, T. J. and Basara, J. B.: Enhancing model-based land surface temperature estimates using multiplatform microwave observations, *J. Geophys. Res. Atmospheres*, 118, 577–591, doi:10.1002/jgrd.50113, 2013a.
- Holmes, T. R. H., Crow, W. T. and Hain, C.: Spatial patterns in timing of the diurnal temperature cycle, *Hydrol. Earth Syst. Sci.*, 17(10), 3695–3706, doi:10.5194/hess-17-3695-2013, 2013b.
- 35 Holmes, T. R. H., Crow, W. T., Hain, C. R., Anderson, M. and Kustas, W. P.: Diurnal temperature cycle as observed by thermal infrared and microwave radiometers., *Remote Sens. Environ.*, 158C, 110–125, doi:10.1016/j.rse.2014.10.031, 2015.

- Jackson, T. J., Le Vine, D. M., Hsu, A. Y., Oldack, A., Starks, P. J., Swift, C. T., Isham, J. D. and Haken, M.: Soil moisture Mapping at regional scales using microwave radiometry: The Southern Great Plains hydrology experiment, *IEEE Trans. Geosci. Remote Sens.*, 37, 2136–2149, 1999.
- 5 Lei, F., Crow, W. T., Shen, H., Parinussa, R. M. and Holmes, T. R. H.: The Impact of Local Acquisition Time on the Accuracy of Microwave Surface Soil Moisture Retrievals over the Contiguous United States, *Remote Sens.*, 7(10), 13448–13465, doi:10.3390/rs71013448, 2015.
- Marek, M. V., Janouš, D., Taufarová, K., Havránková, K., Pavelka, M., Kaplan, V. and Marková, I.: Carbon exchange between ecosystems and atmosphere in the Czech Republic is affected by climate factors, *Environ. Pollut.*, 159(5), 1035–1039, doi:10.1016/j.envpol.2010.11.025, 2011.
- 10 Norman, J. M. and Becker, F.: Terminology in thermal infrared remote sensing of natural surfaces, *Agric. For. Meteorol.*, 77(3–4), 153–166, doi:10.1016/0168-1923(95)02259-Z, 1995.
- Owe, M. and Van de Griend, A. A.: On the relationship between thermodynamic surface temperature and high-frequency (37 GHz) vertically polarized brightness temperature under semiarid conditions., *Int. J. Remote Sens.*, 22, 3521–3532, doi:10.1080/01431160110063788, 2001.
- 15 Owe, M., De Jeu, R. A. M. and Holmes, T. R. H.: Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.-Earth Surf.*, 113, F01002, doi:10.1029/2007JF000769, 2008.
- Parinussa, R., Holmes, T. R. H., Yilmaz, M. T. and Crow, W. T.: The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations, *Hydrol. Earth Syst. Sci.*, 15, 3135–3151, doi:10.5194/hess-15-3135-2011, 2011.
- 20 Prigent, C., Jimenez, C. and Aires, F.: Towards “all weather”, long record, and real-time land surface temperature retrievals from microwave satellite observations, *J. Geophys. Res. Atmospheres* [online] Available from: <http://onlinelibrary.wiley.com/doi/10.1002/2015JD024402/full> (Accessed 17 May 2016), 2016.
- Van Wijk, W. R. and Ubing, D. S.: Radiation, in *Physics of plant environment.*, pp. 62–101, North-Holland Publ. Co., Amsterdam., 1963.

Table 1. List of ground validation targets detailing geographic location and IGBP vegetation type. Local parameters determined for each station are weight (W) in spatial average, long wave emissivity and clear sky absorption-

Cluster ID, name	Site ID	Geographic location	Vegetation (IGBP)	W	ϵ_L median (st.dev)	Clear sky absorbtion (A_{clr})	Notes
A. DE-Thuringia	DE-Hai	51.0792 °N, 10.453 °E	Deciduous Broadleaf Forest	.16	0.993 (0.001)	0.33	(a)
	DE-Geb	51.1001 °N, 10.9143 °E	Croplands	.80	0.983 (0.004)	0.31	
B. DE-Dresden	DE-Tha	50.9636°N, 13.5669°E	Evergreen needleleaf forest	.36	0.983 (0.005)	0.26	
	DE-Kli	50.8928 °N, 13.52250 °E	Croplands	.61	0.993 (0.001)	0.27	
C. CZ-Billy Kris	CZ-BK1	49.50021°N, 18.5368°E	Evergreen needleleaf forest	.72	0.985 (0.001)	0.29	(a)
	CZ-BK2	49.4944°N, 18.5429°E	Grasslands	.23	0.985 (0.001)	0.30	
D. ES-LMa	ES-LMa	39.9415°N, 5.7734 °W	Savannas	.77	0.987 (0.001)	0.25	
E. US-Marys River	US-MRf	44.6465°N, 123.5515°W	Evergreen needleleaf forest	1.0	0.995 (0.000)	0.27	
F. US-Woodward	US-AR1	36.4267°N, 99.42°W	Grasslands	.98	0.993 (0.003)	0.23	
G. US-SGP Main	US-ARM	36.6058 °N, 97.4888°W	Croplands	1.0	0.963 (0.011)	0.23	
H. US-Wind River	US-Wrc	45.8205 °N, 121.9519 °W	Evergreen needleleaf forest	.99	0.993 (0.004)	0.23	
I. US-Santa Rita	US-SRC	31.9083 °N, 110.8395 °W	Open Shrublands	.5	0.960 (0.008)	0.21	
	US-SRM	31.8214 °N, 110.8661 °W	Woody Savannas	.5	0.983 (0.006)	0.21	
J. US-Audubon	US-Aud	31.5907 °N, 110.5092 °W	Grasslands	.99	0.950 (0.003)	0.20	
K. US-Lucky Hills	US-Whs	31.7438 °N, 110.0522 °W	Open Shrublands	1.0	0.972 (0.017)	0.19	
L. US- <u>Woodward</u>	US-AR2	36.6358 °N, 99.5975 °W	Grasslands	1.0	0.992 (0.003)	0.26	
M. US-Kansas	US-KFS	39.0561 °N, 95.1907 °W	Grasslands	1.0	0.945 (0.015)	0.25	

a. Sites are located in neighboring grid cells.
Reference: C: (2011),

Table 2. Percentage coverage for two LST products.

SAT Product	ALL	MSG Domain 2009-2011			
		All	Frost-Free Days		
			All	Clear Sky	Clouded Cloudy Sky
TIR	36	42	47	94	14
MW	50	55	64	75	56

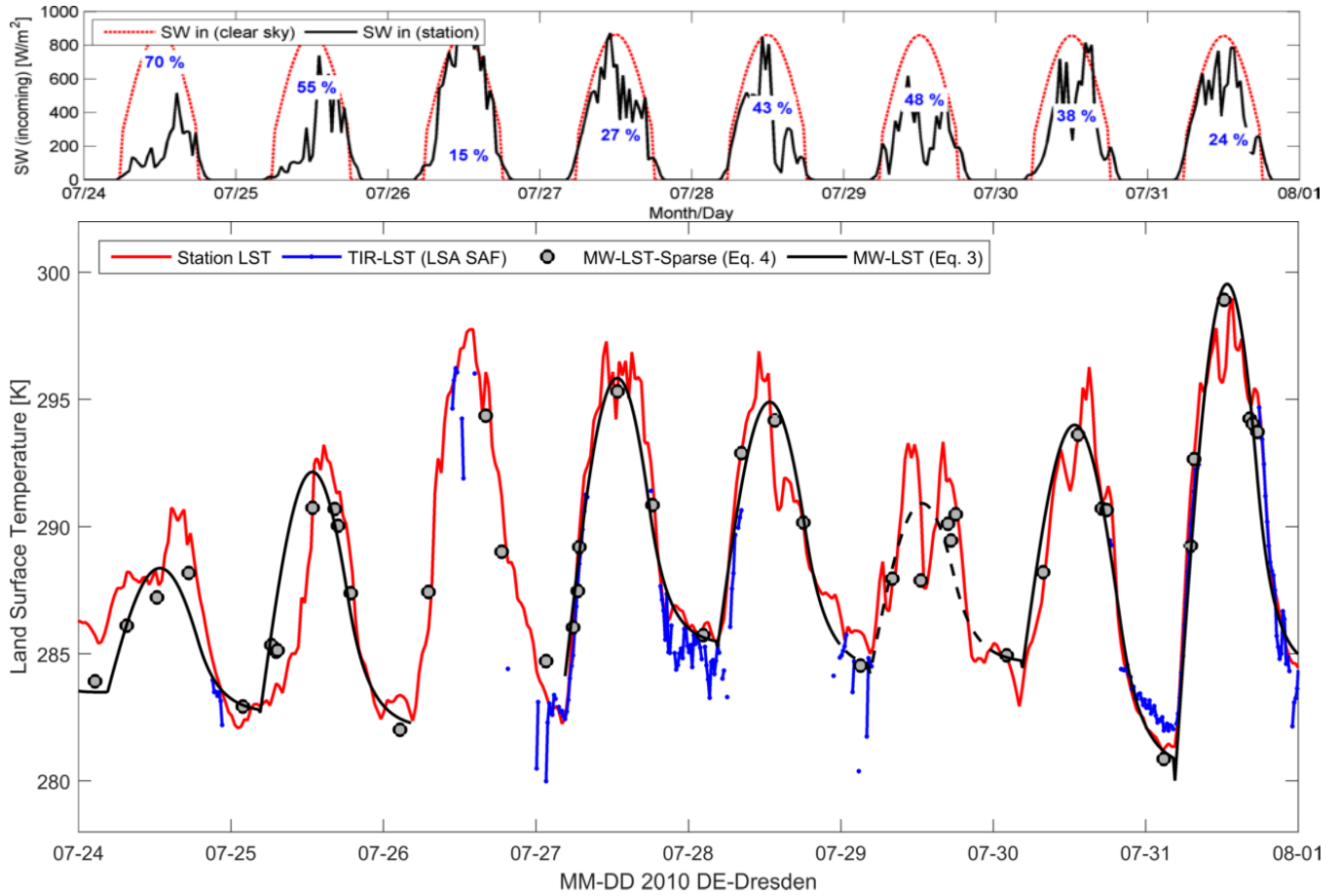
Table 3. Summary of ‘clear-sky’ validation results.

European Fluxnet	Statistic	TIR-LST	MW-LST	MW-LST-Sparse
A: DE-Thuringia	ubRMSE	2.2	2.4	2.4
	BIAS	0.4	1.2	1.1
	N	2050	2636	525
B: DE-Dresden	ubRMSE	1.6	1.7	1.7
	BIAS	-0.6	-0.7	-0.9
	N	4157	4521	1245
C: CZ-BK	ubRMSE	2.6	2.6	2.6
	BIAS	2	2	1.9
	N	2787	2985	850
D: ESLMa	ubRMSE	1.7	2.5	2.6
	BIAS	0.4	0.8	0.7
	N	12055	9652	2560
Multi Site Average	ubRMSE	2.1	2.3	2.3
	 BIAS 	0.9	1.2	1.2

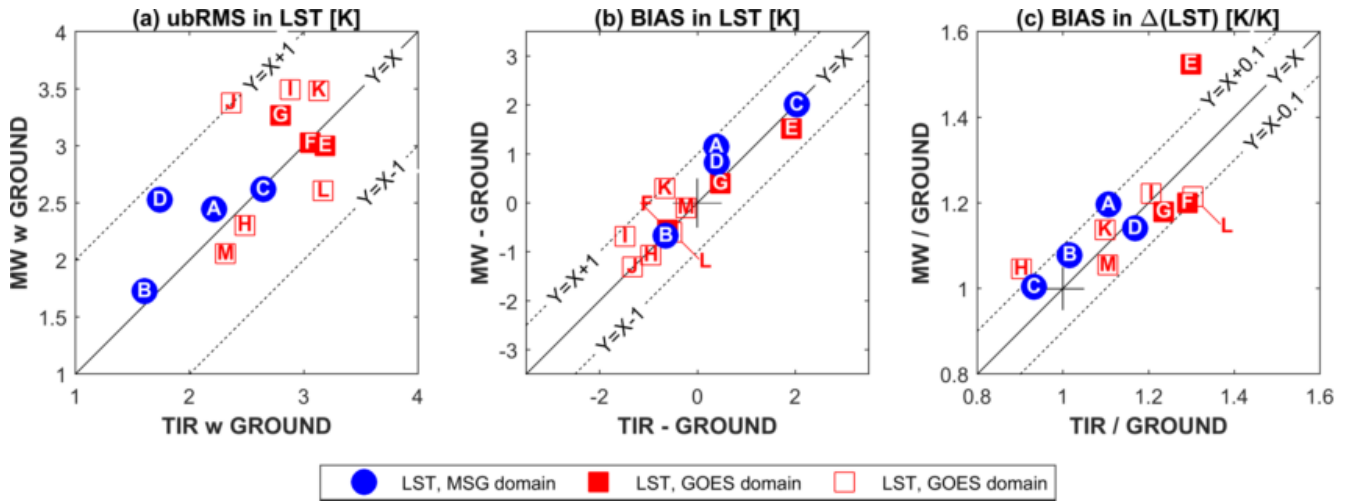
Ameriflux	Statistic	TIR-LST	MW-LST	MW-LST-Sparse
E: US-Marys River	ubRMSE	3.2	3	3.3
	BIAS	1.9	1.5	1.7
	N	2810	1991	756
F: US-Woodward	ubRMSE	3.1	3	2.8
	BIAS	-0.6	-0.5	-0.5
	N	2925	10386	3811
G: US-SGP Main	ubRMSE	2.8	3.3	3
	BIAS	0.5	0.4	0.6
	N	2895	8362	3105
H: US-Wind River	ubRMSE	2.5	2.3	2.5
	BIAS	-0.9	-1.1	-1.1
	N	4219	6915	1687
I: US- Santa Rita	ubRMSE	2.9	3.5	3.3
	BIAS	-1.5	-0.7	-1.2
	N	8908	13478	5312
J: US-Audubon	ubRMSE	2.4	3.4	3.4
	BIAS	-1.3	-1.3	-1.6
	N	3301	8345	3078
K: US-Lucky Hills	ubRMSE	3.1	3.5	3.5
	BIAS	-0.7	0.3	-0.1
	N	7870	10786	4545
L: US-AR2	ubRMSE	3.2	2.6	2.4
	BIAS	-0.5	-0.6	-0.6
	N	2033	8766	3124
M: US-Kansas	ubRMSE	2.3	2.1	2
	BIAS	-0.2	-0.1	-0.4
	N	1454	4788	1627
Multi Site Average	ubRMSE	2.8	3	2.9
	 BIAS 	0.9	0.7	0.9

Table 4. Validation results broken down by MW satellite (all data, descending (D), ascending (A) path), aggregated for the 13 fluxnet sites.

Satellite	Path	2009 Linear regression			2015 diurnal scaling			N
		RMSE	SEE	BIAS	RMSE	SEE	BIAS	
AMSR-E		4.5	3	0.7	3.2	2.6	0.2	6227
	D	4.4	1.9	3	3.1	2.1	-1.1	2817
AMSR2	A	4.6	2.7	-1.5	3.3	2.6	1.4	3410
	D	4.4	2.4	1.6	3	1.9	0.8	843
WindSat	D	5	1.7	3.7	2.6	2	-0.7	347
	A	3.8	1.6	-0.3	3.2	1.6	1.9	478
SSM/I		3.5	2.3	0.8	3.1	2.5	-0.6	3080
	D	3.3	2.1	1.4	3.1	2	-0.2	1617
	A	3.7	1.9	0.2	3	2.1	-0.9	1463
	D	3.8	2.4	0.9	3.1	2.6	-0.3	5401
	A	3.5	2.3	1.1	3.1	2.2	0.2	2740
	A	4	2.1	0.8	3	2.3	-0.8	2661
AVERAGE (All Sites)		4.1	2.4	1.2	3.1	2.4	0.6	
FOREST (Sites: C, E, H)		5.1	2.1	4.3	3	2	0.5	
LOW VEGETATION		3.9	2.5	0.9	3.2	2.5	0.7	



5 Figure 1. An example of the available data at station B, showing 8-day time series of station measured shortwave incoming radiation (S^{\downarrow} , or SW in, top) and LST (bottom graph). In the top graph S^{\downarrow} (black lines) is compared to the clear sky expected value, S^{\downarrow}_{clear} (Eq. 7, red dash), to illustrate the computation of the cloud cover proxy (A_{cloud} , Eq. 8, values in blue text). In the lower graph the station LST is compared to the TIR and MW satellite LST products. The dashed line is an occurrence where the MW-LST is masked due to a high misfit between sparse observations and the diurnal model.



5 Figure 2. Comparison between TIR-LST (X-axis) and MW-LST (Y-axis) in terms of their validation metrics with station LST for frost-free and cloud-free days. From left to right the three panels show (a) ubRMSE, (b) mean bias, and (c) bias in $\Delta(LST)$. Each marker represents the statistics as calculated for individual locations as identified by the letter (see Table 1 for definition). For the GOES domain the filled markers highlight the stations used in the cloud analysis. Black lines provide visual support and indicate targets (eg. 1:1 line, cross at zero bias (b,c)).

10

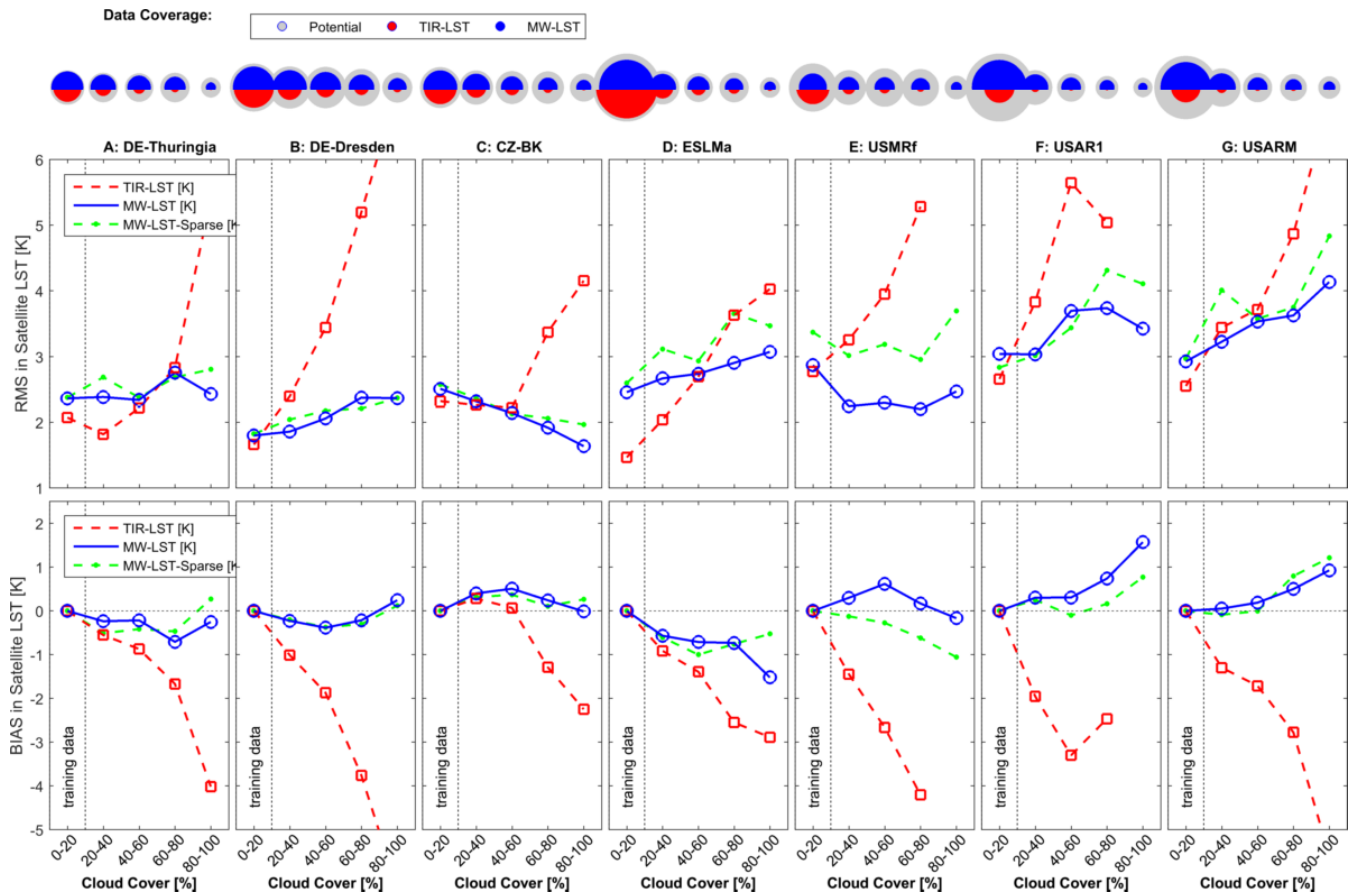


Figure 3. RMSE and bias of satellite LST with regression corrected in situ data for five levels of cloud cover (Acloud, Eq. 8-10). From left to right are locations A-G (see Table 1 for site information). Results for TIR-LST (red) are contrasted with those for MW (blue). Markers indicate that more than 15 days with data were available for a particular cloud cover bin. Green dashed lines indicate the results for MW-LST-Sparse. For each site and cloud interval the percentage coverage of the temporal record is depicted in the top row with half-rounds in proportion to the number of data pairs. The potential number of data pairs (grey) refers to the number of in situ data points for each cloud bin. The actual number of data pairs is superimposed on this for MW (blue) and TIR (red). (MW: blue, TIR: red, and potential coverage in grey).