Demonstrating the Suitability of the Radiance-Based Method for Assessing the Accuracy of MODIS Land Surface Temperature Products

César Coll, Raquel Niclòs, Jesús Puchades, Vicente García-Santos, Martín Perelló, and Lluís Pérez-Planells

***Abstract*— Validation of satellite land surface temperature (LST) products is usually performed through direct comparison with ground-measured LSTs (temperature-based or T-b validation) and with the radiance-based (R-b) method where reference ground LSTs are obtained from at-sensor radiances, atmospheric temperature and water vapor profiles, and emissivity measurements. While T-b is the preferred method, it is only applicable to a few, single-component covers that are thermally homogeneous from the ground measurement scale (a few m) to the satellite scale (usually 1 km). The indirect R-b method is an alternative to extend the LST validation at a global scale over cover types where the T-b method is not feasible. In this study, we used ground LST measurements taken in a thermally homogeneous site coincident with moderate resolution imaging spectroradiometer (MODIS) Terra and Aqua overpasses (86 matchups) to assess the accuracy of the R-b LSTs derived from MODIS data. Mean bias (ground minus R-b LSTs) of −0.1 K and root mean square error (RMSE) of 0.8 K were obtained, showing good performance against ground measurements. Then, we apply operationally the R–b method for validating MODIS Level 2 LST products M*D11 and M*D21 over six test sites comprising a varied range of surfaces (4267 cases). The emissivity values necessary for the R-b calculations were obtained from 1) the M*D11 and M*D21 emissivity product and 2) an independently modeled emissivity using ground measurements and vegetation cover fraction estimates. For the M*D11 product, we obtained an overall bias (product minus R-b LST) of −0.5 K and RMSE of 0.9 K with either the product or the modeled emissivities. However, a significant difference was found between daytime and nighttime bias (−1.0 and 0.0 K, respectively). The daytime cold bias of M*D11 was attributed to an ill-tuning of the algorithm for high surface temperature and atmospheric humidity conditions. For the M*D21 product, the overall bias (RMSE) was 0.2 K (0.6 K) with the product emissivity, and 0.6 K (1.1 K) with the modeled emissivity, both in nighttime and daytime. These results compare well with recent studies and contribute to the global assessment of MODIS LST uncertainty.**

Manuscript received 12 April 2024; revised 26 June 2024; accepted 26 August 2024. Date of publication 4 September 2024; date of current version 17 September 2024. This work was supported in part by the Generalitat Valenciana under Project PROMETEO/2021/016 and in part by the Spanish Ministry of Science and Innovation and the Spanish State Research Agency (MCIN/AEI/10.13039/501100011033) under Project PID2020-118797RBI00 (Tool4Extreme). *(Corresponding author: César Coll.)*

César Coll, Raquel Niclòs, Jesús Puchades, Vicente García-Santos, and Martín Perelló are with the Department of Earth Physics and Thermodynamics, Faculty of Physics, University of Valencia, 46100 Burjassot, Spain (e-mail: cesar.coll@uv.es; raquel.niclos@uv.es; jesus.puchades@uv.es; Vicente.Garcia-Santos@uv.es; Martin.Perello@uv.es).

Lluís Pérez-Planells is with IMK-ASF, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany (e-mail: Lluis.Perez@uv.es).

Digital Object Identifier 10.1109/TGRS.2024.3454377

***Index Terms*— Emissivity, land surface temperature (LST), M*D11 and M*D21 products, radiance-based (R-b) validation.**

I. Introduction

LAND surface temperature (LST) is a key magnitude for many of Earth's sciences. The Global Climate Observing System and the European Space Agency's Climate Change Initiative program have recognized LST as an Essential Climate Variable for the monitoring of the Earth's climate [1], [2]. The availability of global, highly repetitive, long-term, and consistent LST data provided by thermal infrared (TIR) satellite sensors has increased their use in different applications in recent years, including estimation of evapotranspiration and soil moisture, monitoring of agricultural drought, urban heat island, thermal anomalies, and climate change indexes [3]. One of the key points conditioning the use of satellite LSTs is the necessity of a consistent assessment of the LST uncertainty. In addition, continuous evaluation of satellite LST estimates may help to improve and refine existing LST retrieval algorithms.

The committee on Earth observation satellites (CEOSs) compiled a best practice protocol for satellite LST validation [4] where two main validation methods are generally considered: The temperature-based (T-b) and the radiance-based (R-b) methods. T-b is the preferred validation method [4] and consists of a direct comparison of satellite LST with concurrent ground LST measurements in thermally homogenous test sites. The T-b method requires ground TIR radiometers calibrated against traceable standards [5], in situ measurements of surface emissivity and downwelling sky radiance for emissivity correction of radiometric temperatures, and especially that the ground measurements, which correspond to a small field of view (typically few m^2), represent the satellite-scale LST (usually 1 km^2).

Lagouarde et al. [6], [7] showed that atmospheric turbulence might induce high-frequency temporal fluctuations of surface temperature at the ground instrument scale even for single-component homogeneous surfaces. They proposed frequent sampling using 2–3 instruments per site at distances enough to avoid correlation in surface temperatures and averaging during a few minutes around the satellite overpass to reduce the uncertainty in the ground-measured LSTs. Moreover, a comprehensive uncertainty assessment of the ground

LSTs must be provided [8], showing the significance of the ground data with regard to the satellite LSTs. The uncertainty of ground LST measurements includes calibration and emissivity correction errors, plus spatial-temporal variability of LST within the satellite pixel area at the acquisition time, which can only be accounted for by multiinstrument and frequent sampling and it is usually the largest source of error. The above issues make the collection of ground LSTs a highly demanding task in terms of resources. It is limited to a few cover types, with single-component temperature, spatially uniform from a few meters to 1 km such as lakes, dense vegetation, and bare surfaces [9], [10], [11]. This limits the applicability of the T-b method for global scale validation of satellite LSTs.

As an alternative, the R-b method [12] does not rely on ground LST measurements. Instead, it uses the brightness temperatures in two bands in the 10–12.5 μm atmospheric window, coincident atmospheric profiles of pressure, temperature, and water vapor, and emissivity in the considered bands to simulate the LST through radiative transfer modeling. Strictly, the R-b method cannot be considered as a direct, independent validation in comparison with the T-b method. However, it can be a useful alternative when ground measurements of LST are not viable. The R-b method is affected by several sources of uncertainty, especially the atmospheric profiles used for radiative transfer simulation and the emissivity values. Such uncertainties can be assessed through the difference between LSTs derived in two bands of the 10–12.5 μm window (δ), which should be close to zero in ideal cases. Therefore, we can define a δ threshold value to filter out cases with excessive error [12]. Since emissivity in the 10–12.5 μm window shows small variations among most natural land covers, sites with uniform emissivity are more frequent than thermally homogeneous sites, so the R-b method can be potentially applied to a larger number of sites than the T-b method.

In this article, we focused on the moderate resolution imaging spectroradiometer (MODIS) onboard Terra and Aqua satellites, which currently provides two global, long-term, four times a day LST and emissivity products at 1 km resolution: M*D11_L2 and M*D21_L2 (here M*D11 and M*D21, with MOD for Terra and MYD for Aqua), Collection C6.1. M*D11 is probably the most extensively used and validated satellite LST product and is usually considered a reference for intercomparison with other satellite LST products [3]. The more recent M*D21 LST product has been validated in comparatively fewer studies. The CEOS LST validation protocol [4] defined four validation stages. Stage 1 means that only a small (<30) set of sites and times have been evaluated. Stage 2 implies that a significant set of locations and times have been evaluated as well as the spatial and temporal consistency against similar products at a global scale. Stage 3 is achieved when uncertainties are well quantified and statistically characterized in a robust way over global conditions. Finally, stage 4 requires that stage 3 be updated for new versions and time series. According to the MODIS data web page (https://modis.gsfc.nasa.gov/), the M*D11 LST product is currently in validation stage 2, while M*D21 was recently promoted from stages 1 to 2. Thus, it is worthwhile pursuing the validation of the M*D11 and M*D21 LST products for different biomes and atmospheric conditions in order to promote the validation stage of the products. In order to accomplish this goal, the R-b validation method appears as the most suitable and operational possibility for global LST validation.

The objective of this article is twofold. First, demonstrating the feasibility of the R-b method applied to MODIS data through a comparison with ground LSTs measured in a homogeneous site in order to assess the uncertainty of the method and the threshold for the δ difference. To our knowledge, this is the first study where T-b and R-b LSTs are compared in an experimental test site. Second, showing the operational application of the R-b method for M*D11 and M*D21 LST validation in different test sites encompassing various surface types and atmospheric conditions during a two-year period. Novelties of the article include: 1) a simplified formulation of the R-b method and the δ difference; 2) validation of R-b LSTs with ground-measured LSTs in a thermally homogeneous site; and 3) operational R-b validation of M*D11 and M*D21 LST products, and evaluation of the impact of emissivity inputs.

The article continues with a description of the simplified R-b method and the datasets used in the present study. Section III shows the results of the comparison of R-b LSTs against ground-measured LSTs and the operational validation of M*D11 and M*D21 LSTs. Section IV discusses the results obtained and Section V gives the main conclusions of the study.

II. Method and Materials

A. Simplified R-b Method

The R-b method [12] starts from the top-of-atmosphere (TOA) thermal radiance measured in band i of a satellite sensor, L_i^{sen}, which can be written as

$$L_i^{\text{sen}} = \left[\varepsilon_i B_i\left(T_g\right) + (1-\varepsilon_i)L_i^{\text{sky}}\right]\tau_i(\theta) + L_i^{\text{atm}}(\theta) \quad (1)$$

where subindex i denotes a band-averaged magnitude according to the band's response function, ε_i is the emissivity, T_g is the LST, B_i is the Planck function for blackbody radiance, L_i^{sky} is the downwelling sky irradiance divided by π, $\tau_i(\theta)$, and $L_i^{\text{atm}}(\theta)$ are respectively the atmospheric transmittance and upwelling radiance at the zenith angle θ. If surface emissivity is known and atmospheric transmittance and radiances can be calculated from atmospheric profiles and a radiative transfer model, the reference LST can be obtained from (1) according to

$$T_{ig} = B_i^{-1}\left[\frac{L_i^{\text{sen}} - L_i^{\text{atm}}}{\varepsilon_i\tau_i(\theta)} - \frac{1-\varepsilon_i}{\varepsilon_i}L_i^{\text{sky}}\right] \quad (2)$$

where B_i^{-1} is the inverse Planck function. The LST obtained from (2) is represented by T_{ig} since it may depend on the band used for the retrieval. Typically, the R-b method is applied separately to two bands splitting the 10–12.5 μm atmospheric window (e.g., MODIS bands 31 and 32), hereafter referred to as bands $i = 1$ and 2, respectively.

In an ideal case, T_{ig} should be identical in both bands but they are affected by uncertainties in the emissivity and atmospheric profiles. For most natural surfaces such as water,

vegetation, and soils, emissivity in the 10–12.5 μm window are close to unity and show small variability among surface types [13], [14], [15]. Therefore, emissivity uncertainties can be kept within reasonable limits. More critical is the need for accurate concurrent atmospheric profiles. The accuracy of the radiative transfer calculations can be assessed through the difference in the LST obtained from the two window bands, $\delta = T_{1g} - T_{2g}$. Atmospheric effects in the window region are mainly due to the continuum absorption of water vapor, so they are highly correlated between bands and they are larger at 12 μm than at 11 μm. So, when the atmospheric profile used is overcorrecting the real atmospheric effect in band 1, that is, the difference between the actual and derived LST, $T_g - T_{1g}$, is negative, the overcorrection in T_{2g} will be larger and then $\delta < 0$. The reverse is true in the case of under correction in band 1 ($T_g - T_{1g} > 0$), which will yield to a larger underestimation in band 2 and thus $\delta > 0$. Therefore, the difference $\delta = T_{1g} - T_{2g}$ should be within a small range around zero when atmospheric profiles are accurate and uncertainties in surface emissivities are small.

The δ threshold that determines the valid cases can be obtained through an uncertainty analysis of the various factors involved in the derivation of T_{ig}, so a range of δ values can be defined in order to keep the uncertainty in T_{1g} within certain limits [12], [16]. In this article, we adopted a different strategy using ground-measured LSTs for validating the derived temperatures and establishing the valid δ range.

The simplified R-b method can be summarized in three steps as follows.

1) Using the at-sensor radiances in bands 1 and 2, atmospheric parameters, and emissivity data, the surface temperature is obtained for both bands, T_{1g} and T_{2g} (2).
2) The difference $\delta = T_{1g} - T_{2g}$ is calculated and only cases for which δ is within a small interval around zero are selected as valid cases.
3) For the valid cases, the calculated temperature in band 1, T_{1g}, is considered the reference LST for satellite LST validation. Band 1 is selected because atmospheric correction errors are smaller than in band 2.

In the original formulation [12], δ was defined as the brightness temperature difference ($T_1 - T_2$) obtained from the actual satellite data minus the value calculated from direct (upward) simulation using T_{1g} as input LST in (1). The new δ definition seems more straightforward and relies on the same rationale behind the original definition. In addition, we used an approximation for the band-averaged Planck function, B_i, as an analytical function of temperature (e.g., [8]) so T_{ig} can be readily inverted from (2) and the iterative procedure to calculate T_{ig} [12] is not necessary.

Fig. 1 shows a flow diagram of the methodology applied in this study. Sections II-B–II-D describe the different datasets used.

B. MODIS Data and Atmospheric Profiles

MODIS products were downloaded from https://modis.gsfc.nasa.gov/data/ and include M*D21KM (Level 1B calibrated TOA radiances), M*D03 (Geolocation), and M*D11 and M*D21 LST products, all at 1 km resolution. The M*D11 product uses the generalized split-window method [17] with the brightness temperatures in bands 31 (10.8–11.3 μm) and 32 (11.8–12.3 μm) and a priori classification-based band emissivities. The M*D21 product [18], [19] uses a physics-based temperature-emissivity separation algorithm [20] with the three MODIS TIR bands 29 (8.4–8.7 μm), 31 and 32. The M*D21 algorithm simultaneously retrieves LST and emissivities in the three bands involved. Both products provide the LST, the band emissivities, and a theoretical estimation of the LST uncertainty on a pixel basis.

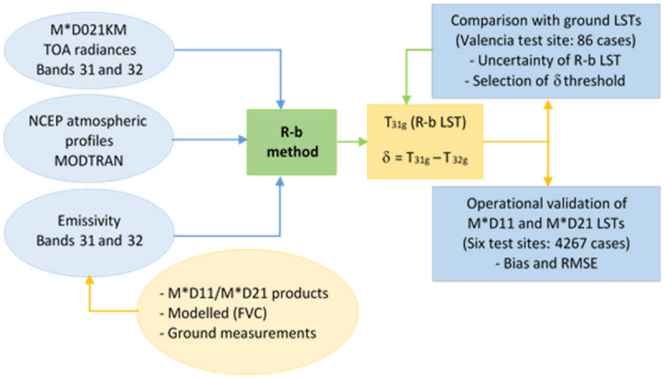


Fig. 1. Flow diagram showing the steps of the methodology used.

TABLE I
TEST SITES FOR R-b VALIDATION, SHOWING THE LOCATION, NUMBER OF SCENES, COMPONENT EMISSIVITIES, AND RANGE OF FVC. *WET BARE SOIL

			vegetation		soil		
Site	Longitude, Latitude	scenes	band 31	band 32	band 31	band 32	FVC range
Lake	39°20'24''N, 0°21'00''W	380	N/A	N/A	N/A	N/A	N/A
Rice field	39°16'12'', 0°19'12''W	397	0.985	0.980	0.970 0.972*	0.972 0.977*	0.03-1
Shrubland	39°13'30"N, 0°53'42"W	811	0.985	0.986	0.972	0.977	0.16-0.33
Forest	41°35'31"N, 2°33'00"W	775	0.977	0.977	0.972	0.977	0.27-0.53
Vineyard	39°34'48"N, 1°16'12"W	792	0.972	0.973	0.967	0.959	0.03-0.20
Olive grove	37°49'30"N, 3°58'12"W	1112	0.976	0.976	0.967	0.959	0.09-0.21

Two different MODIS datasets were used in this study. First, a limited dataset corresponds to a large and homogeneous rice field in Valencia, Spain where simultaneous ground LST measurements were performed (see Section II-D). The dataset includes 65 Terra cases and 21 Aqua cases from 2002 to 2022, all daytime. They were used to compare the R-b LSTs against independent T-b LSTs to assess the uncertainty of the method and select the δ threshold. Second, an extended dataset covering six selected test sites (see Table I) in the years 2020 and 2021 was used for operational R-b validation with a total of 4267 cases with approximately the same number for Terra and Aqua, daytime, and nighttime in each site. We checked that all pixels covering the test sites were classified as cloud-free and flagged with the highest quality in the auxiliary data.

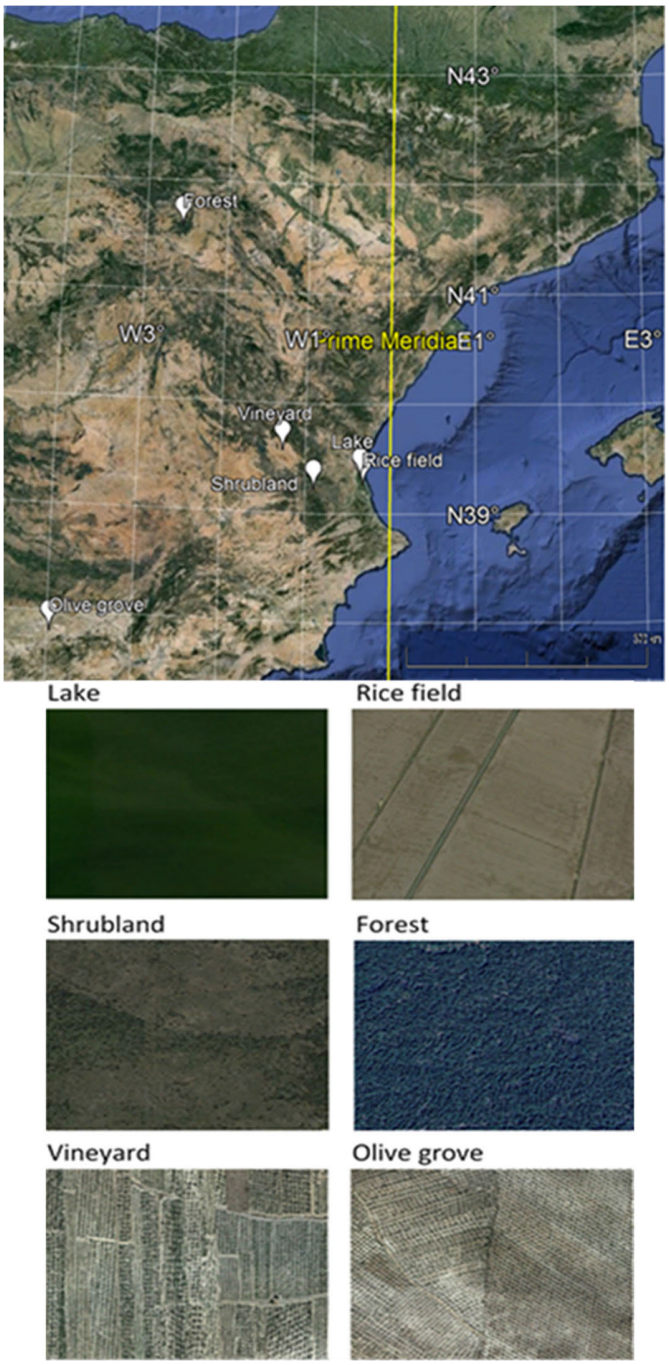


Fig. 2. (Top) Location and (bottom) detailed views of the test sites. The rice field is shown in conditions of dry bare soil.

The test sites are located within large uniform areas containing at least 5 × 5 pixels (3 × 3 for the lake site). Fig. 2 shows the locations of the sites and detailed views of the different covers. For each site and scene, we extracted the four pixels closest to the coordinates of the test site to obtain the satellite value by weighted spatial interpolation.

While the test sites of Table I are restricted to a limited geographical area, they encompass a significant variety of surface covers, from water and nearly full vegetation conditions to partial vegetation covers and practically bare soils. As an indicator of the variety of conditions included, brightness temperatures in band 31 ranged from approximately 280–320 K in daytime, and from 270 to 290 K in nighttime.

We used atmospheric profiles from the National Centers for Environmental Prediction (NCEP) global model. It provides atmospheric profiles over a 1° × 1° grid at 00:00, 06:00, 12:00, and 18:00 UTC, which were spatially and temporally interpolated to the coordinates of each site and the time of sensor overpass. The atmospheric profiles were input in the radiative transfer code MODTRAN 5 [21] to obtain spectral transmittances and radiances. As NCEP profiles only arrive up to 29 km over ground level, upper atmospheric levels were added from the mid-latitude summer and winter standard atmospheres, depending on the date. We added the vertical distribution of CO_2, O_3, and other minor gases included in the MODTRAN code. The atmospheric spectral transmittances and radiances were integrated into the MODIS bands 31 and 32 using the corresponding response functions. The total column water vapor (W) ranged from 0.3 to 3.5 cm in the study period, showing a wide range of mid-latitude atmospheric conditions.

C. Emissivity Data

An important issue of the R-b method is the emissivity inputs required to derive the reference LSTs. For the operational R-b validation, we have considered two options: 1) the emissivity data provided by each LST product, that is, a priori classification-based emissivity for M*D11 and the physics-based derived emissivity for M*D21 and 2) a dynamic emissivity database defined for each site taking into account the nature of the site and available ground or laboratory emissivity measurements for water, vegetation, and soil components.

For partially vegetated surfaces, emissivities were modeled with the vegetation cover method (VCM, [22]) which relies on the fractional vegetation cover (FVC). Since MODIS does not provide an operational FVC product, we used the Level 2 LST product of Sentinel-3, which includes FVC at 1 km as an auxiliary dataset. Sentinel-3 FVC estimates are based on vegetation indexes obtained from the near-infrared bands (0.66 and 0.87 μm) of the Sea and LST Radiometer. FVC values are provided with a temporal resolution of ten days acquired from a moving temporal window of 30-day composites. The FVC values for the MODIS observation dates were obtained by linear interpolation from the closest cloud-free Sentinel-3 dates, usually around three days before the MODIS overpass. Table I shows the approximate range of FVC for the land sites.

The product emissivities are closely related to the LSTs provided by each product, either because they are used as inputs for LST derivation in the M*D11 product or calculated simultaneously with LST in the M*D21 product. Thus, such emissivity data may be considered a favorable case for R-b validation. Since the modeled emissivities are independent of the products, they are useful to check the impact of emissivity on the validation.

Water emissivity and its angular dependence are well known (e.g., [23]), and it is spatially uniform and constant in time.

In the lake site, we applied the parametrization of [24] to calculate water emissivity as a function of observation zenith angle from 0° to 65° in bands 31 and 32. The rice field site is the same as used for ground measurements (Section II-D) showing different covers through the year from water, wet and dry bare soil, and rice crops in various stages with FVC ranging roughly from 0 to 1. For this site, we measured emissivity values in the field for full vegetation and wet and dry bare soil conditions (Table I) using the box method and the TES method applied to multiband field radiometer CIMEL CE-312 [25] as described in [26]. Emissivity measurements were repeated during the campaigns, Table I shows the average emissivity values at nadir for CE-312 bands B3 (10.2–11.0 μm) and B2 (10.9–11.7 μm), which are comparable to MODIS bands 31 and 32, respectively. For water-flooding conditions, we modeled the emissivity in the same way as for the lake site. Field emissivity measurements are not usually available for most land sites, so we can consider the rice field site as an ideal case for R-b validation.

The shrubland and forest are natural evergreen vegetation showing seasonal variability in FVC, with maximum values of 0.3 for the shrubland in spring and 0.5 for the forest in summer, and minimum values over 0.2. The vineyard and olive grove are dryland crops, with minimum FVC values lower than 0.1 and not exceeding 0.2 during the year, so they can be considered representative of arid or semi-arid conditions. For these sites, the band i emissivity ε_i was modeled from VCM according to [15]

$$\varepsilon_i = \varepsilon_{vi} f + \varepsilon_{si}(1-f) + 4(-0.435\varepsilon_{si} + 0.4343)(1-f)f \tag{3}$$

where f is FVC and ε_{vi} (ε_{si}) is the vegetation (soil) component emissivity. Component emissivities are shown in Table I for each site. Emissivity values for the vineyard and shrubland sites were obtained from ground measurements [15], [27], whereas they were taken from [28] for the olive grove sites and from [29] for the forest. Given the emissivity and FVC values of the different test sites (Table I), we have estimated that an uncertainty of 0.1 in FVC leads to an average uncertainty of 0.005 in emissivity according to (3).

The emissivity model of (3) is simple and easy to apply, only depending on FVC and vegetation and soil emissivities that are considered static for each site during the year. However, leaf emissivities depend on the growing conditions, usually decreasing from fresh to dry and senescent stages. Emissivity variations during the growing season depend on the vegetal species and are difficult to know, so they have not been considered in the model. Vegetation emissivities of Table I correspond to full development conditions so the resulting modeled emissivities can be overestimated in some sites during dry and senescent periods.

Fig. 3 shows a comparison of the MOD11, MOD21, and the modeled emissivity values in band 31 (similar results were obtained for Aqua and band 32). For the lake site, MOD11 and modeled emissivities only depend on the zenith angle and agree very well up to 40°; however, MOD11 appears to overestimate emissivity for larger zenith angles. For the land sites, the MOD11 and modeled emissivities show small variability during the year. The modeled emissivity takes into account the different stages of the rice field site (water, bare soil, and vegetation) and changes in FVC for the other land sites resulting in relatively small seasonal variability with maximum values in spring-summer, while MOD11 shows mostly constant emissivity values for each site. In general, the agreement between MOD11 and the modeled emissivity is reasonable for sites with medium to high vegetation cover (rice fields in summer, shrubland, and forest), the largest differences being found for the vineyard and olive grove sites with small FVC. In these cases, MOD11 emissivities are higher and close to vegetation values, whereas the modeled emissivities are closer to those of bare soil due to the low FVC. On the other hand, modeled emissivities for the evergreen shrubland site are somewhat higher than MOD11 emissivities, which are closer to bare soil.

Oppositely, MOD21 emissivities show values smaller than the MOD11 and modeled emissivity data and a large variability with no correlation with the zenith angle in the lake site nor with day of year in the land sites. The reason for the large variability is that MOD21 emissivities are physically retrieved together with LST so they vary from case to case as they are affected by several sources of uncertainty such as instrumental noise, atmospheric correction, and algorithm error [19]. Therefore, we also show in Fig. 3(a) the smoothing of the daily MOD21 emissivities using a moving average over intervals of 5° in viewing angle for the lake site and of ten days for the land sites in order to match the temporal resolution of the FVC data used for the modeled emissivities. According to Fig. 3, the MOD21 emissivity variability appears to be larger in spring-summer (larger atmospheric humidity) and the maximum values are usually close to MOD11 and modeled emissivities. The smoothed emissivities show better consistency with regard to the other emissivity databases and a clearer seasonal dependence than the daily MOD21 product.

The impact of the emissivity inputs on the R-b method can be assessed through a sensitivity analysis of (2). The uncertainty in T_{1g} corresponding to an uncertainty of 0.005 in ε_1 is 0.37 K for the average atmospheric conditions of the dataset. Assuming the same uncertainty for ε_2 and that emissivity uncertainties are uncorrelated in both bands, the impact of emissivity uncertainty alone on δ is 0.52 K. This shows that emissivity is a crucial input, especially, for the calculation of δ, and that R-b validation results depend on the emissivity database used.

D. Ground LST Measurements

The rice field site has been used as a T-b validation site for different TIR sensors since 2002 [10], [16], [26], [27], [30], [31], [32], [33]. In order to assess the accuracy of the R-b derived LSTs with regard to independent ground data, we used 65 cases of ground measurements concurrent with Terra in 2002–2022, and 21 cases concurrent with Aqua in 2016–2022. All cases correspond to daytime, cloud-free conditions with overpass time between 10:10 and 11:40 UTC for Terra and 12:40 and 13:45 UTC for Aqua. Most of the data correspond to full vegetation cover conditions (75%), but there are a few cases for water (13%) and bare soil (12%).

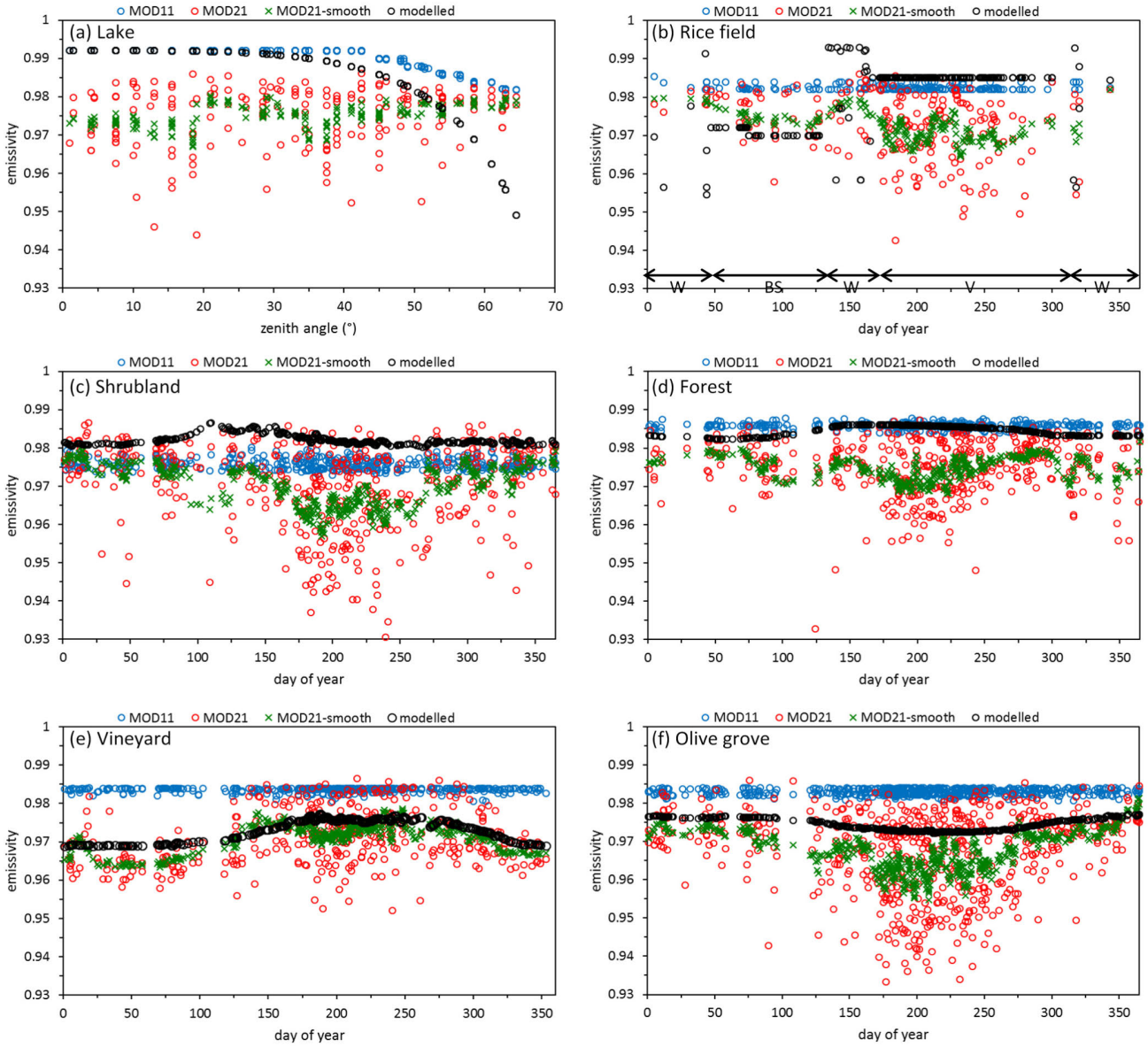


Fig. 3. Emissivity in MODIS band 31 from MOD11 (blue), MOD21 (red), and MOD21 smoothed with moving average (green) and modeled (black) for (a) water as a function of zenith observation angle, (b) rice field with water (W), bare soil (BS) and vegetation (V) stages indicated, (c) shrubland, (d) forest, (e) vineyard, and (f) olive grove as a function of the day of year.

Ground LST measurements were performed using two to four portable radiometers CE-312 [25]. Instruments were calibrated against absolute standards in the TIR instrumentation comparisons held by CEOS in the National Physical Laboratory (NPL) in 2009, 2016 [8], and more recently in 2022 [5]. According to these results, the calibration uncertainty of the CE-312 radiometers was better than 0.20 K in the 273–323 K range. In the NPL comparisons, we also checked the accuracy of the Landcal P80P blackbody (https://www.ametek-land.com), which showed good stability and an error smaller than 0.15 K compared with the NPL standard. The Landcal P80P blackbody source is used for CE-312 calibration in our laboratory on a yearly basis, so the calibration of the radiometers can be regularly updated and reliably estimated for each measurement campaign.

Field measurements were made close to the nadir, so we checked that the corresponding MODIS zenith observation angle was lower than 40° in order to avoid thermal anisotropy effects [6]. The field of view of CE-312 is 10°, which corresponds to a sampling area of about 30 cm in diameter. To assess the variability of the ground temperatures within the satellite pixel, radiometers were assigned to different parts of the test site and carried on along predetermined transects taking measurements each 5–10 m and at a rate of five measurements/min approximately [10], [32].

We corrected radiometric temperatures for surface emissivity effects using ground measurements of emissivity (Table I) and downwelling sky radiance. The final ground LST was calculated as the mean value of the individual LSTs taken by all ground radiometers during 5 min around the satellite

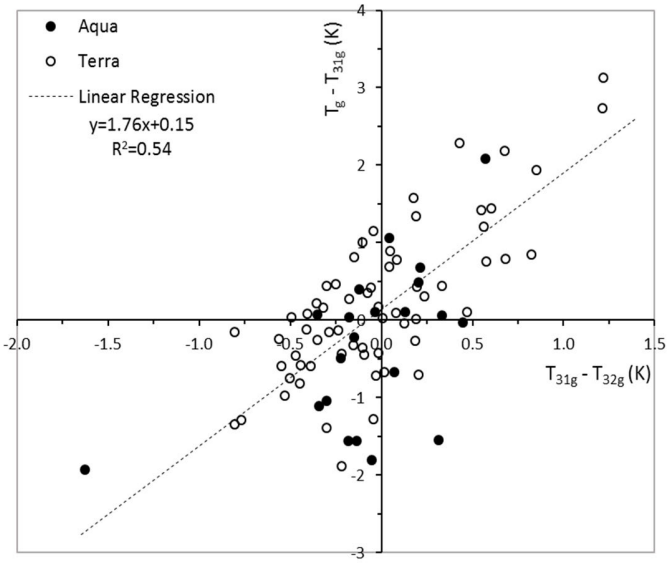


Fig. 4. Difference between measured and R-b surface temperatures in band 31 against the difference in R-b surface temperatures in bands 31 and 32. Terra (Aqua) cases are marked with open (closed) circles. The linear regression and determination coefficient (R^2) are shown.

overpass. The uncertainty of the ground LSTs was estimated considering the following sources of error: 1) calibration of ground radiometers; 2) emissivity measurements; 3) downwelling sky radiance measurements; and 4) spatial-temporal LST variability. The latter was taken as the standard deviation of the LSTs measured by the different radiometers during the 5 min period. Sources (1–3) typically accounted for an error of 0.5 K [8], while the total uncertainty in the ground-measured LST dataset ranged from 0.7 to 1.4 K, the largest source being the spatial and temporal variability.

III. Results

A. Comparison of R-b and Ground-Measured LSTs

We applied the R-b method to the dataset with concurrent ground LST measurements to compare both LSTs. Fig. 4 plots the difference between the ground-measured LSTs (T_g) and R-b reference LSTs (T_{31g}) against the difference $\delta = T_{31g} - T_{32g}$ for all available cases. It shows a good correlation between both magnitudes, with large negative (positive) δ values usually corresponding to large negative (positive) values in $T_g - T_{31g}$, while we have small $T_g - T_{31g}$ values for δ around 0. The linear regression in Fig. 4 yields an acceptable determination coefficient ($R^2 = 0.54$) and an intercept close to zero (0.15 K) as expected. The slope of the linear regression is close to 2, which suggests that the range in δ should be ± 0.5 K to keep the LST uncertainty within ± 1.0 K (comparable with ground measurements). This threshold coincides with that considered by [19] and [34] from theoretical sensitivity analysis.

In Fig. 4, a few points with δ values close to zero have negative values of $T_g - T_{31g}$ in excess of -1.5 K. Such cases (1 Terra and 4 Aqua) correspond to summer conditions (warm and wet atmosphere), with ground LST uncertainties between 1.0 and 1.2 K, so the large differences between the ground and R-b LSTs may be compatible with the combined uncertainty of $T_g - T_{31g}$.

Taking only the cases meeting -0.5 K $\leq \delta \leq 0.5$ K, the mean value or bias and root mean square error (RMSE) of $T_g - T_{31g}$ were respectively 0.0 and 0.8 K for Terra (48 valid cases), -0.4 and 0.9 K for Aqua (19 valid cases), and -0.1 and 0.8 K overall. These results show that the accuracy that can be obtained with R-b LSTs is comparable to that of ground-measured LSTs. Therefore, the R-b ground temperatures can be used as an alternative to T-b LSTs for validation when δ is within ± 0.5 K.

B. Application of the R-b Method for Operational Validation

The R-b method was applied to the extended dataset for operational validation of MODIS LST products. Figs. 5 (MOD11) and 6 (MOD21) plot the difference between the product LST (T_{MOD11} or T_{MOD21}) and the R-b LST (T_{31g}) against the difference $\delta = T_{31g} - T_{32g}$ for each test site (daytime and nighttime data combined). In each case, we show the results obtained from the product emissivity and the modeled emissivity. Only Terra data are presented, results for Aqua were mostly similar. Figs. 5 and 6 show a similar relationship between the temperature differences as displayed in Fig. 4. The best correlations were obtained for MOD21 using the product emissivities, where coefficients of determination (R^2) were in the 0.7–0.8 range (Fig. 6). However, they were lower (0.2–0.5) when we used the modeled emissivity data. For the MOD11 product (Fig. 5), R^2 values ranged roughly between 0.3 and 0.5 depending on the site, with a small difference between the product emissivity and the modeled emissivity.

Taking into account only the cases passing the δ condition, we calculated the mean bias and RMSE of the product minus R-b LST differences for M*D11 (Table II) and M*D21 (Table III). Results are shown for each site and all sites combined, and for daytime, nighttime, and all together. Terra and Aqua data are considered separately. Results obtained when using the product and the modeled emissivities are shown.

Tables II and III also show the fraction (in %) of valid cases passing the δ condition for each site and instrument (MOD and MYD) with the product and modeled emissivities. In general, valid cases were more than 70% for all sites and emissivity databases. The only exceptions were the vineyard and olive grove sites with the modeled emissivities (24%–56%). The lower number of valid cases in such sites may be due to the dissimilarity between modeled and product emissivities, especially with regard to M*D11 [Figs. 3(e) and (f), 5(e) and (f), and 6(e) and (f)]. This means that the validation database could be significantly different for the vineyard and olive grove sites when comparing the different emissivity databases, which may affect the validation results. However, the validation databases contain more than 100 data points per site and instrument even in the worst case. It is also noticeable that nighttime datasets generally had a larger fraction of valid cases than daytime datasets, the most favorable being MYD at nighttime.

In Sections III-C and III-D, we describe the validation results of Tables II and III.

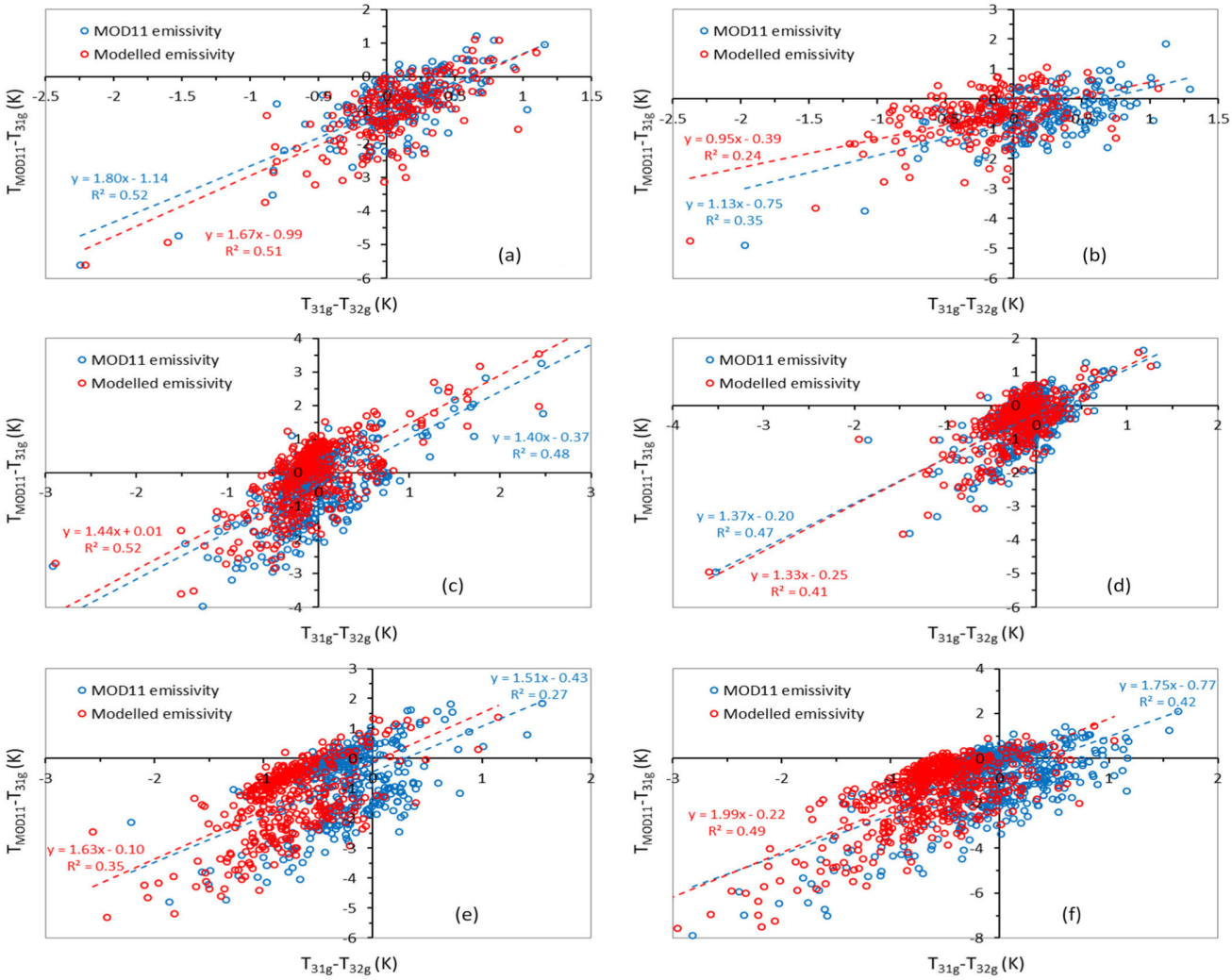


Fig. 5. Difference between MOD11 and R-b LSTs in band 31 against the difference $\delta = T_{31g} - T_{32g}$ for the six test sites when the MOD11 emissivity data (blue) and the modeled emissivity values (red) are used. The linear regression and determination coefficient (R^2) are shown for each case. (a) Lake. (b) Rice field. (c) Shrubland. (d) Forest. (e) Vineyard. (f) Olive grove.

C. M*D11 Validation

Table II shows a good consistency between Terra and Aqua M*D11 LSTs both in the daytime and nighttime. Generally, Aqua provided slightly better results than Terra, but differences in bias and RMSE were below 0.2 K in absolute value for most of the sites, with maximum differences of 0.3 K for the shrubland site and usually within 0.1 K for the other sites. Validation results were mostly similar with either the product or the modeled emissivities, as anticipated in Fig. 5. Considering the results from the product emissivity for all sites, daytime, nighttime, and Terra and Aqua combined, the overall bias was −0.5 K with RMSE of 0.9 K. Per site, biases varied from −0.2 K for the forest to −0.8 K for the lake, while RMSEs ranged from 0.6 K for the forest to 1.1 K for the olive grove. Considering the daytime data only, we found negative biases (product colder than R-b LST) of −1.0 K for all sites combined, ranging between −0.5 K for the forest and −1.4 K for the olive grove. However, the overall bias was 0.0 K for nighttime data, the extreme case being −0.4 K for the lake site and ranging between −0.1 and 0.2 K for the other land sites. The different daytime and nighttime biases of M*D11 can be noticed in Fig. 5, where two clusters of data corresponding to the two conditions are clearly visible, especially for the vineyard and olive grove sites.

D. M*D21 Validation

According to Table III, Terra and Aqua results were very similar as well for the M*D21 using the product emissivity data, with biases and RMSEs within ±0.1 K on average for all sites, in daytime and nighttime. When the modeled emissivities were used, Terra-Aqua differences were slightly larger in some cases, especially for the vineyard and the olive grove at nighttime. It may be because such cases had the lowest number of data points passing the δ threshold as mentioned in Section III-B. However, Terra-Aqua differences were still within ±0.3 K on average. On the other hand, the impact of the emissivity database was larger for the M*D21 product than for M*D11, with the lowest biases and RMSEs being obtained from the product emissivities.

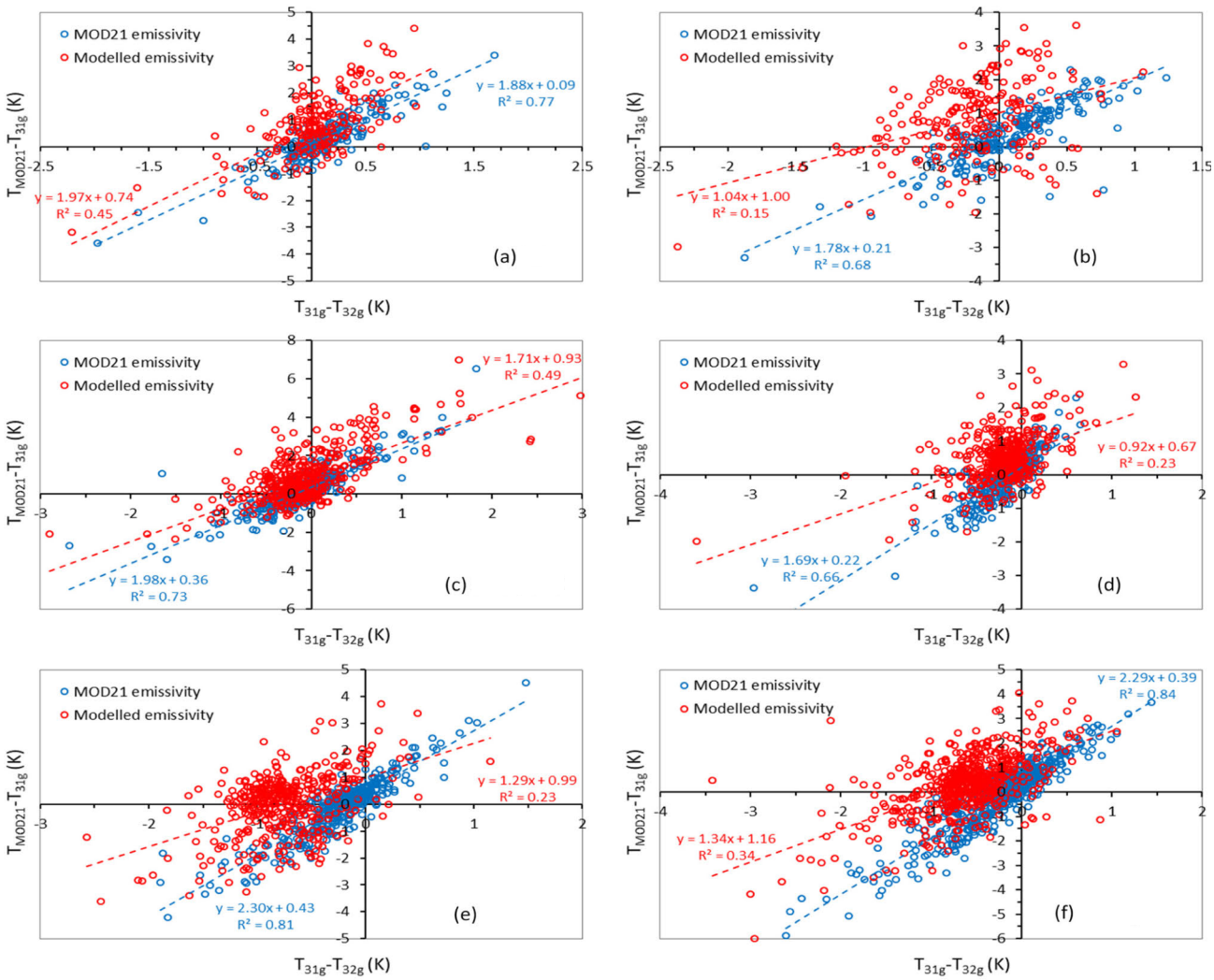


Fig. 6. Difference between MOD21 and R-b LSTs in band 31 against the difference $\delta = T_{31g} - T_{32g}$ for the six test sites when the MOD21 emissivity data (blue) and the modeled emissivity values (red) are used. The linear regression and determination coefficient (R^2) are shown for each case. (a) Lake. (b) Rice field. (c) Shrubland. (d) Forest. (e) Vineyard. (f) Olive grove.

Considering the results with the product emissivities for all sites, all acquisition times and platforms combined, the overall bias was 0.2 K with RMSE of 0.6 K. Considering the sites separately, biases ranged from 0.1 K for the shrubland and forest sites to 0.3 K for the rice field site, while RMSEs ranged between 0.4 K for the forest site and 0.7 K for the lake, rice field and olive grove site. Unlike M*D11, differences between nighttime and daytime results were not significant, the daytime (nighttime) bias being 0.1 K (0.3 K) for all sites, Terra and Aqua combined. When using the modeled emissivities, we can see a uniform increase of about 0.6 K in bias and RMSEs with regard to the product emissivities for most sites.

IV. Discussion

A. Comparison With LST Product Uncertainty

The M*D11 and M*D21 products provide an estimate of the LST uncertainty for each LST value. An example is shown for the olive grove site in Fig. 7(a), where the LST uncertainties of MOD11 and MOD21 LSTs (Terra only) are plotted against the day of the year. The M*D11 LST uncertainty was uniform with a value around 0.5 K for all test sites, daytime and nighttime, Terra, and Aqua data. The M*D21 LST uncertainty was typically larger, with minimum values around 1 K and high variability at the central days of the year where the uncertainty may reach over 3 K, and an average value of 1.3 K for all sites, acquisition times, and platforms.

In Fig. 7(b), the product LST uncertainties are plotted against $W/\cos\theta$, which represents the atmospheric water vapor content in the satellite-viewing path. We can see that the MOD21 LST uncertainty increases clearly with $W/\cos\theta$ since the uncertainties in the atmospheric correction process required for the M*D21 product increase for high atmospheric water vapor loads and surface temperatures, as is the case for summer and early autumn. For the M*D11 uncertainty, there is only a small increase from 0.4 to 0.7 K with atmospheric humidity.

The RMSEs found for M*D11 (Section III-C) are compatible with the product uncertainty for the nighttime data only, while they are larger for daytime due to the relatively large

TABLE II

MEAN BIAS AND RMSE (ALL IN K) OF THE M*D11 PRODUCT MINUS R-b LSTS FOR EACH SITE AND ALL SITES COMBINED WITH THE PRODUCT EMISSIVITY AND THE MODELED EMISSIVITY, FOR DAYTIME, NIGHTTIME AND ALL COMBINED, AND FOR TERRA (MOD) AND AQUA (MYD). THE FRACTION (IN %) OF MOD AND MYD VALID CASES IN EACH SITE IS GIVEN IN PARENTHESES

		M*D11 emissivity		modelled emissivity	
lake		MOD (85)	MYD (84)	MOD (83)	MYD (86)
day	bias	-1.1	-0.8	-1.3	-1.0
	RMSE	1.3	1.0	1.5	1.1
night	bias	-0.5	-0.4	-0.7	-0.5
	RMSE	0.7	0.7	0.8	0.8
all	bias	-0.9	-0.7	-1.1	-0.8
	RMSE	1.1	0.9	1.3	1.0
rice field		MOD (72)	MYD (72)	MOD (77)	MYD (81)
day	bias	-1.0	-0.9	-0.7	-0.6
	RMSE	1.1	1.0	1.0	1.0
night	bias	-0.2	0.0	0.0	-0.1
	RMSE	0.5	0.4	0.5	0.6
all	bias	-0.6	-0.5	-0.5	-0.4
	RMSE	0.9	0.8	0.8	0.9
shrubland		MOD (81)	MYD (84)	MOD (80)	MYD (85)
day	bias	-0.9	-0.6	-0.5	-0.2
	RMSE	1.2	0.9	0.9	0.7
night	bias	0.2	0.1	0.5	0.4
	RMSE	0.4	0.4	0.6	0.6
all	bias	-0.4	-0.2	-0.1	0.1
	RMSE	0.9	0.7	0.8	0.6
forest		MOD (87)	MYD (95)	MOD (85)	MYD (94)
day	bias	-0.6	-0.5	-0.6	-0.5
	RMSE	0.9	0.7	0.8	0.7
night	bias	0.0	0.1	-0.1	0.1
	RMSE	0.4	0.3	0.4	0.3
all	bias	-0.3	-0.1	-0.4	-0.2
	RMSE	0.7	0.5	0.6	0.5
vineyard		MOD (79)	MYD (88)	MOD (24)	MYD (35)
day	bias	-1.4	-1.1	-1.5	-1.1
	RMSE	1.5	1.4	1.6	1.3
night	bias	0.1	0.2	0.3	0.2
	RMSE	0.5	0.5	0.7	0.4
all	bias	-0.6	-0.3	-0.7	-0.5
	RMSE	1.1	0.9	1.3	1.0
olive grove		MOD (76)	MYD (74)	MOD (38)	MYD (56)
day	bias	-1.5	-1.3	-1.5	-1.3
	RMSE	1.7	1.6	1.7	1.5
night	bias	0.0	-0.1	-0.2	-0.3
	RMSE	0.4	0.4	0.4	0.5
all	bias	-0.7	-0.5	-0.8	-0.9
	RMSE	1.2	1.0	1.2	1.1
all sites		MOD (80)	MYD (83)	MOD (60)	MYD (69)
day	bias	-1.1	-0.9	-1.0	-0.8
	RMSE	1.3	1.1	1.3	1.1
night	bias	0.0	0.0	0.0	0.0
	RMSE	0.4	0.4	0.5	0.5
all	bias	-0.6	-0.4	-0.6	-0.4
	RMSE	1.0	0.8	1.0	0.9

TABLE III

MEAN BIAS AND RMSE (ALL IN K) OF THE M*D21 PRODUCT MINUS R-b LSTS FOR EACH SITE AND ALL SITES COMBINED WITH THE PRODUCT EMISSIVITY AND THE MODELED EMISSIVITY, FOR DAYTIME, NIGHTTIME AND ALL COMBINED, AND FOR TERRA (MOD) AND AQUA (MYD). THE FRACTION (IN %) OF MOD AND MYD VALID CASES IN EACH SITE IS GIVEN IN PARENTHESES

		M*D21 emissivity		modelled emissivity	
lake		MOD (79)	MYD (78)	MOD (83)	MYD (86)
day	bias	0.2	0.3	0.8	1.0
	RMSE	0.6	0.8	1.2	1.4
night	bias	0.2	0.3	0.9	0.7
	RMSE	0.6	0.7	1.2	1.0
all	bias	0.2	0.3	0.8	0.9
	RMSE	0.6	0.7	1.2	1.3
rice field		MOD (81)	MYD (80)	MOD (77)	MYD (81)
day	bias	0.2	0.3	1.0	1.1
	RMSE	0.7	0.8	1.6	1.7
night	bias	0.4	0.4	0.8	0.6
	RMSE	0.6	0.5	1.2	0.9
all	bias	0.3	0.3	0.9	0.9
	RMSE	0.7	0.7	1.4	1.5
shrubland		MOD (85)	MYD (90)	MOD (80)	MYD (85)
day	bias	0.2	0.3	1.0	1.0
	RMSE	0.7	0.6	1.4	1.3
night	bias	0.0	0.0	0.2	0.3
	RMSE	0.5	0.3	0.6	0.6
all	bias	0.1	0.1	0.7	0.6
	RMSE	0.6	0.5	1.1	1.0
forest		MOD (89)	MYD (98)	MOD (85)	MYD (94)
day	bias	0.0	0.1	0.5	0.9
	RMSE	0.5	0.4	0.9	1.1
night	bias	0.2	0.1	0.5	0.4
	RMSE	0.4	0.2	0.7	0.5
all	bias	0.1	0.1	0.5	0.6
	RMSE	0.5	0.3	0.8	0.8
vineyard		MOD (80)	MYD (91)	MOD (24)	MYD (35)
day	bias	-0.2	0.1	-0.1	0.1
	RMSE	0.6	0.7	1.1	1.0
night	bias	0.4	0.3	1.3	0.5
	RMSE	0.6	0.4	1.6	0.8
all	bias	0.1	0.2	0.5	0.3
	RMSE	0.6	0.5	1.4	0.9
olive grove		MOD (75)	MYD (87)	MOD (38)	MYD (56)
day	bias	-0.1	0.0	0.7	0.6
	RMSE	0.7	0.8	1.3	1.2
night	bias	0.5	0.4	0.9	0.4
	RMSE	0.7	0.5	1.3	0.7
all	bias	0.3	0.2	0.8	0.5
	RMSE	0.7	0.6	1.3	1.0
all sites		MOD (81)	MYD (89)	MOD (60)	MYD (69)
day	bias	0.0	0.2	0.6	0.8
	RMSE	0.6	0.7	1.2	1.2
night	bias	0.3	0.2	0.8	0.5
	RMSE	0.6	0.4	1.1	0.7
all	bias	0.2	0.2	0.7	0.6
	RMSE	0.6	0.5	1.2	1.0

cold biases obtained in most sites. For the M*D21 product, the RMSEs obtained here with both the product and the modeled emissivity are smaller than the product uncertainty for most sites in daytime and nighttime.

B. Impact of Emissivity on R-b Validation

For M*D11 (Table II), validation results were slightly better for the product emissivities than for the modeled emissivities, but the difference was rather small, with changes within ± 0.2 K in biases and ± 0.1 K in RMSEs for most of the sites, all acquisition times, and platforms. This is because the two emissivity databases were similarly defined through classification and show comparable values (Fig. 3). The impact of emissivity was larger for M*D21, the best validation results being obtained from the product emissivity (overall RMSE of 0.6 K). This shows the consistency of the LST and emissivities

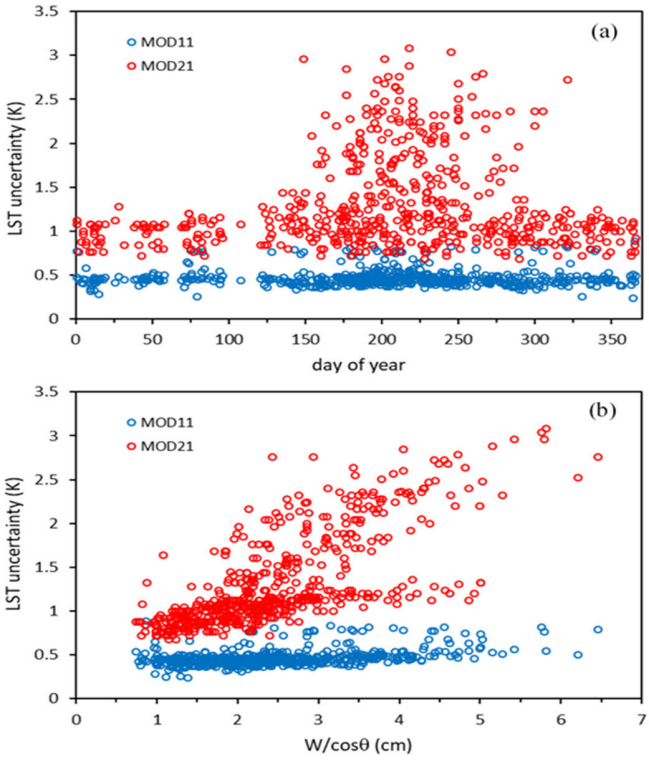


Fig. 7. Uncertainty of the LST products MOD11 (blue) and MOD21 (red) for the olive grove site as a function of (a) day of the year and (b) water vapor content in the satellite viewing path $W/\cos\theta$. All data points available for the site are included.

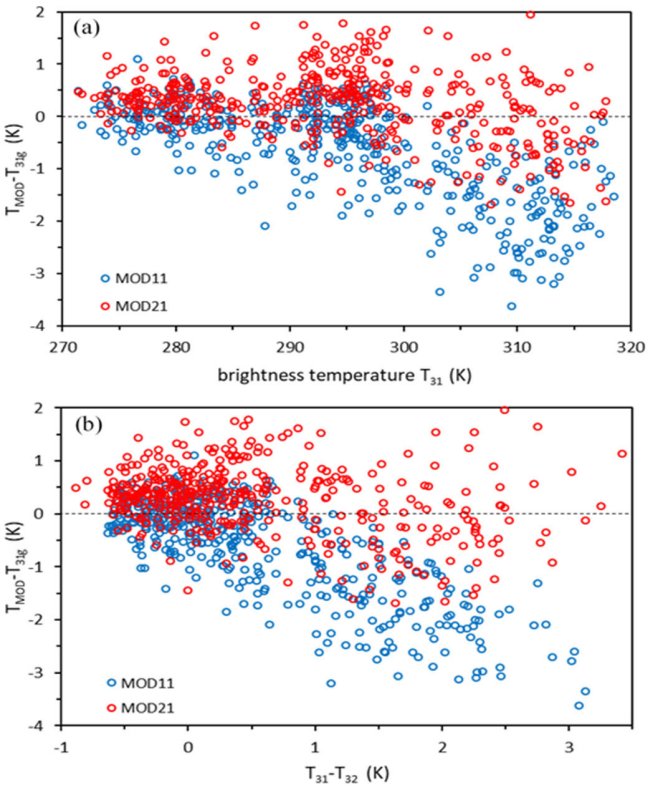


Fig. 8. Difference between the MODIS LST product and the R-b reference temperature T_{31g} for MOD11 (blue) and MOD21 (red) in the olive grove site as a function of (a) MODIS brightness temperature in band 31 and (b) brightness temperature difference $T_{31} - T_{32}$. Only data points passing the δ condition are included.

of the M*D21 product, which are simultaneously obtained from the physics-based TES method. As noted before, the bias and RMSE increased when using the independently modeled emissivities. For the overall dataset, the bias was 0.6 K and RMSE was 1.1 K, still showing an acceptable performance. Per site, biases ranged from 0.4 K (vineyard) to 0.9 K (rice field) and RMSEs from 0.8 K (forest) to 1.4 K (rice field). The increase in bias is explained since, according to Fig. 3, the modeled emissivities are typically larger than M*D21 emissivities, thus yielding smaller R-b LSTs. As with the product emissivities, no significant differences were found between daytime and nighttime biases.

C. M*D11 Daytime Bias

Section III-C shows a cold daytime bias of about −1.0 K on average for M*D11. The daytime bias may be related to the underestimation of product LSTs at high temperatures, as shown in Fig. 8(a) where the difference $T_{\mathrm{MOD11}} - T_{31g}$ (with T_{31g} obtained from the product emissivities) is plotted against the brightness temperature T_{31} for the olive grove site. Results for the other sites and MYD11 data were comparable. We can see that the negative bias occurs for brightness temperatures higher than 295 K, which correspond mostly to daytime. In Fig. 8(b) the differences $T_{\mathrm{MOD11}} - T_{31g}$ are plotted against the brightness temperature differences $T_{31} - T_{32}$, which play an important role in the M*D11 split-window algorithm. Typically, large values of $T_{31} - T_{32}$ correspond to daytime cases in summer. According to Fig. 8(b), the underestimation of T_{MOD11} appears to increase linearly with the brightness temperature difference. This may indicate an ill-tuning of the split-window coefficients in the M*D11 algorithm for high surface temperatures and atmospheric water vapor contents.

Water emissivity is uniform, constant, and well-known, so the modeled emissivities should be close to the real value. However, daytime and nighttime biases were considerable (−1 and −0.5 K, respectively in Table II). Part of the negative bias could be due to the overestimation of M*D11 emissivity for viewing angles higher than 40° shown in Fig. 3(a). Taking into account only the cases with $\theta < 40°$, the daytime bias decreased to −0.8 K, while the nighttime bias increased to −0.7 K. The similar negative biases can be due to water temperatures being in the 295–300 K range for more than 50% of both daytime and nighttime cases analyzed here.

Oppositely, the M*D21 product showed no significant difference in daytime and nighttime biases. In Fig. 8, we plotted as well the differences $T_{\mathrm{MOD21}} - T_{31g}$ against T_{31} and $T_{31} - T_{32}$ for the olive grove site showing no correlation. Dispersion increased for higher values of T_{31} and $T_{31} - T_{32}$, which typically correspond to spring and summer with humid atmospheres, in concordance with the M*D21 product uncertainty patterns shown in Fig. 7. Results were mostly similar for the other sites and MYD data.

D. Comparison of R-b Validation Results With Previous Studies

Wan [35] validated the product M*D11 using the R-b method and found biases mostly within ±0.6 K for several desert sites in North Africa for daytime, nighttime, Terra, and Aqua data. While daytime cases correspond to high temperatures (up to 320 K), low atmospheric humidity conditions make the brightness temperature difference $T_{31} - T_{32}$ negative (see [35, Tables 3 and 4]) so LST biases should be close to zero according to Fig. 8(b). Duan et al. [36] applied the R-b method to M*D11 data showing mean biases between −0.4 and 0.7 K and RMSEs smaller than 1.0 K. They found similar daytime and nighttime biases, oppositely to the results shown here. R-b validation of MYD11 over sand dunes in China [34] showed negative biases in daytime and nighttime (−1.3 and −0.7 K, respectively for all sites combined). However, biases varied from site to site, with daytime (nighttime) values about −3 and −4 K (−1 and −3 K) for three sites and better than −1.0 K (−0.5 K) for the other sites. The large cold bias of MYD11 LST was attributed to an overestimation of emissivity for the three sites in the MYD11 product. However, emissivity in MODIS bands 31 and 32 shows small variability among different types of soils [13], [14], [15] and cannot totally account for the large observed biases. Perez-Planells et al. [15] analyzed MYD11 daytime LSTs for shrubland and vineyard sites using different sets of emissivity inputs and showed a similar temperature-depending bias as in Fig. 8(a), regardless of emissivity. In addition, the difference between nighttime and daytime bias in [34] is similar to that reported here and may be related to high temperature and atmospheric humidity conditions prevailing in daytime.

In [19], the R-b method was applied to the M*D21 product over desert dunes, vegetated surfaces, and water, using emissivity data from ground measurements. They showed biases within ±0.9 K and RMSEs from 0.5 to 1.5 K depending on the site for daytime and nighttime data combined. More recently, [34] showed R-b results for MYD21 over sand dunes using ground-measured emissivities. Biases and RMSEs were quite consistent among sites and between daytime and nighttime, with a mean bias of 0.5 K (0.7 K) and RMSE of 1.1 K (0.9 K) for all daytime (nighttime) cases. Dune sites in desert atmospheric conditions are likely an optimal case for the MOD21 algorithm because of the high spectral contrast of sand emissivity between bands 31–32 and 29, and the low impact of atmospheric water vapor [19]. The validation sites of the article include cases of vegetated and water surfaces with near gray-body spectrum in humid atmospheres, which are the less favorable cases for M*D21. Nevertheless, the results shown in Table III with the modeled emissivities are comparable to those of [19] and [34] for all sites considered.

E. Comparison of R-b and T-b Validation

The R-b results obtained here compare well with previous T-b validation results of [19] with daytime and nighttime ground measurements in Lake Tahoe and Salton Sea. For MYD11, they found a bias (RMSE) of 0.2 K (0.5 K) in Lake Tahoe (low-temperature range) and −1.1 K (1.5 K) in Salton Sea (high-temperature range). For MYD21, results were 0.7 K (0.9 K) and 0.2 K (1.2 K), respectively. Such error figures are comparable as well to the R-b validation results of [19] mentioned earlier. In the same way, [30] showed T-b validation for MOD11 and MOD21 daytime data over the same rice field site used here for R-b validation, with bias (RMSE) of 0.1 K (0.6 K) for MOD11 and 0.5 (0.6 K) for MOD21.

Several studies on T-b validation of satellite LSTs rely on ground measurements performed by single, tower-mounted pyrgeometers (flux-meters) in meteorological stations instead of radiometers (directional radiance-meters) [34], [37], [38]. Pyrgeometer measurements represent a broadband, hemispherically integrated value while satellite measurements are narrow-band and of directional nature, depending on the zenith and azimuth viewing and solar angles. Therefore, they are not necessarily comparable, especially in the case of sparse vegetation areas. Moreover, climatological stations are not usually located in thermal homogeneous areas so it is not straightforward to assign the ground-measured LST to the whole satellite pixel.

The effect of thermal heterogeneity at the subpixel scale was analyzed in [38] through the standard deviation of 11 × 11 pixels of 90-m ASTER temperatures coincident with a MODIS pixel. However, single-point ground measurements correspond to areas of a few meters and there may be still a large heterogeneity from the ground scale to the 90-m scale. Daytime results showed M*D11 RMSEs larger than 2 K for most sites and biases from −0.1 to −4.7 K, while nighttime RMSEs were <2 K for all sites. Duan et al. [38] attributed the big daytime LST inaccuracies to the thermal heterogeneity of the sites. The selected cases where the ASTER LST standard deviation was <1 K provided the best results, with close to zero bias and RMSE < 1.3 K. For comparison, typical daytime standard deviations of ASTER LSTs for the rice field site used here were about 0.5 K [10].

Li et al. [34] showed T-b validation for MYD11 and MYD21 LSTs using ground measurements in four desert sites carried out by single, mast-mounted instruments with no information provided about the thermal heterogeneity at the sub-pixel scale. Daytime MYD11 LSTs showed large cold biases close to −3 K and RMSE close to 4 K for the four sites on average. For nighttime MYD11 LSTs, the average bias reduced to −1.4 K and the RMSE to 1.7 K. Li et al. [34] attributed such cold biases to emissivity problems with MYD11 in barren surfaces, where they found MYD11 emissivities overestimating MYD21 emissivities by 0.01. While this may account for a part of the nighttime bias, it cannot explain alone the large daytime bias and RMSEs observed. Such errors include a considerable component due to sub-pixel thermal heterogeneity, possible algorithm problems with high temperatures mentioned before, and probably thermal directional effects from comparing nadir ground measurements with satellite measurements at different zenith angles. The T-based validation of MYD21 showed daytime bias of −0.5 K and RMSE of 2.5 K and nighttime bias of 0.2 K and RMSE of 1.0 K. We can see that the daytime bias was reduced considerably, but still, the RMSE is large and may contain heterogeneity and anisotropy issues. The nighttime MYD21 results were

optimum and only comparable to the R-b validation results of [34] discussed in Section IV-D.

V. Conclusion

The R-b method was used for the validation of M*D11 and M*D21 LST products. First, we applied the R-b method to a set of Terra and Aqua daytime data to derive the R-b reference LSTs that were compared with a dataset of coincident ground-measured LSTs in a homogeneous rice field. The comparison showed an agreement between R-b and T-b LSTs better than 1.0 K in RMSE (i.e., similar to typical uncertainty in ground measurements) when the δ difference was within ± 0.5 K. Then, we applied the R-b method to six test sites including water, homogeneous rice paddy, and partially vegetated areas with FVC ranging from nearly zero to 0.5. All sites were at mid-latitudes but still covering a wide range of LST (270–320 K), and atmospheric humidity (0.5–3.5 cm). For each LST product, two emissivity datasets were used for R-b calculations: the corresponding product emissivity and the modeled emissivity using component emissivity measurements and FVC.

According to the results shown here, daytime M*D11 LSTs showed an average bias of -1.0 K (product LST colder than reference LST) and RMSE of 1.2 K for all sites, Terra and Aqua data combined, with either the product or the modeled emissivities. The daytime cold bias was attributed to the unfitting of the MOD11 LST algorithm for high temperatures and atmospheric humidities. For nighttime data, the overall M*D11 LST bias was 0.0 K (ranging from -0.4 to 0.2 K among sites) and RMSE was 0.5 K. M*D21 LSTs showed great stability among sites and between daytime and nighttime, with daytime (nighttime) bias of 0.1 K (0.3 K) and RMSE of 0.7 K (0.5 K) for all sites, Terra and Aqua data combined, when the product emissivities were used. For the modeled emissivities, bias, and RMSE increased consistently for all sites, times of day, and platforms, but still provided acceptable overall bias (0.6 K) and RMSE (1.1 K).

For a meaningful validation of LST products, the uncertainty of the LSTs used as a reference (either T-b or R-b) must be fully characterized and reliably estimated. In the R-b method, uncertainties mainly include atmospheric correction errors coming from inaccurate atmospheric profiles, radiative transfer model errors, and emissivity input uncertainties. Emissivity has an important impact, especially on the δ calculations. For bare surfaces (rock, soil, sand), emissivity and its spectral variation within the 10–12.5 μm band depend on many factors such as composition, moisture, and texture, so uncertainties may be larger than for vegetated or partially vegetated surfaces. Atmospheric correction errors also have an effect on the R-b LSTs, which tend to be larger for warm, humid atmospheres, and high surface temperatures. However, the impact of the error sources can be limited to a useful range through the δ condition. In the T-b method, sources of error include calibration of ground radiometers, emissivity measurement, and most importantly, the LST spatial-temporal variability within the pixel scale. The sub-pixel LST variability can only be fully assessed through multispatial and frequent temporal sampling in single-component thermal homogeneous areas. On the other hand, ground LST measurements are usually performed at the nadir while satellite viewing zenith angles may exceed 60°, so there may be directional effects, especially for partially vegetated surfaces. These issues limit the application of the T-b method, especially in the daytime.

The T-b method must be considered as the reference, independent validation method for satellite LSTs, so long-term and accurate ground-LST measurements in highly homogeneous test sites are necessary. However, the R-b method provides a valuable alternative for most worldwide surface covers and atmospheric conditions, and the accuracy of the R-b derived LSTs can be checked against ground-measured LSTs in selected test sites as in the present article. However, more studies are required for different ground covers in warm and humid tropical environments where atmospheric effects may be larger thus being a challenging case for satellite LST retrieval.

References

[1] *The 2022 GCOS ECVs Requirements (GCOS 245)*, GCOS, Geneva, Switzerland, 2022, pp. 1–244.

[2] R. Hollmann et al., "The ESA climate change initiative: Satellite data records for essential climate variables," *Bull. Amer. Meteorolog. Soc.*, vol. 94, no. 10, pp. 1541–1552, Oct. 2013, doi: 10.1175/BAMS-D-11-00254.1.

[3] Z. Li et al., "Satellite remote sensing of global Land Surface Temperature: Definition, methods, products, and applications," *Rev. Geophys.*, vol. 61, no. 1, pp. 1–77, Mar. 2023, doi: 10.1029/2022rg000777.

[4] P. Guillevic et al., "Land Surface Temperature product validation best practice protocol. Version 1.1," in *Good Practices for Satellite-Derived Land Product Validation*, P. Guillevic, F. Göttsche, J. Nickeson, and M. Román, Eds., Land Product Validation Subgroup (WGCV/CEOS), 2018, p. 58, doi: 10.5067/doc/ceoswgcv/lpv/lst.001.

[5] Y. Yamada et al., "CEOS international thermal infrared radiometer comparison: Part I: Laboratory comparison of radiometers and blackbodies," *J. Atmos. Oceanic Technol.*, vol. 41, pp. 295–307, Mar. 2024.

[6] J.-P. Lagouarde, M. Irvine, and S. Dupont, "Atmospheric turbulence induced errors on measurements of surface temperature from space," *Remote Sens. Environ.*, vol. 168, pp. 40–53, Oct. 2015, doi: 10.1016/j.rse.2015.06.018.

[7] J.-P. Lagouarde, M. Irvine, and P. Guillevic, "New insights of ground-based Land Surface Temperature measurements protocols for improving validation of thermal infrared satellite data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 9184–9187, doi: 10.1109/IGARSS.2019.8898312.

[8] C. Coll et al., "Laboratory calibration and field measurement of Land Surface Temperature and emissivity using thermal infrared multiband radiometers," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 78, pp. 227–239, Jun. 2019, doi: 10.1016/j.jag.2019.02.002.

[9] S. J. Hook, R. G. Vaughan, H. Tonooka, and S. G. Schladow, "Absolute radiometric in-flight validation of mid infrared and thermal infrared data from ASTER and MODIS on the Terra spacecraft using the lake tahoe, CA/NV, USA, automated validation site," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1798–1807, Jun. 2007, doi: 10.1109/TGRS.2007.894564.

[10] C. Coll et al., "Ground measurements for the validation of Land Surface Temperatures derived from AATSR and MODIS data," *Remote Sens. Environ.*, vol. 97, no. 3, pp. 288–300, Jun. 2005, doi: 10.1016/j.rse.2005.05.007.

[11] F.-M. Göttsche, F.-S. Olesen, I. Trigo, A. Bork-Unkelbach, and M. Martin, "Long term validation of Land Surface Temperature retrieved from MSG/SEVIRI with continuous in-situ measurements in Africa," *Remote Sens.*, vol. 8, no. 5, p. 410, May 2016, doi: 10.3390/rs8050410.

[12] Z. Wan and Z.-L. Li, "Radiance-based validation of the V5 MODIS land-surface temperature product," *Int. J. Remote Sens.*, vol. 29, nos. 17–18, pp. 5373–5395, Sep. 2008, doi: 10.1080/01431160802036565.

[13] W. C. Snyder, Z. Wan, Y. Zhang, and Y.-Z. Feng, "Classification-based emissivity for Land Surface Temperature measurement from space," *Int. J. Remote Sens.*, vol. 19, no. 14, pp. 2753–2774, Jan. 1998, doi: 10.1080/014311698214497.

[14] A. C. T. Pinheiro, J. L. Privette, R. Mahoney, and C. J. Tucker, "Directional effects in a daily AVHRR Land Surface Temperature dataset over Africa," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1941–1954, Sep. 2004, doi: 10.1109/TGRS.2004.831886.

[15] L. Pérez-Planells, R. Niclòs, E. Valor, and F.-M. Göttsche, "Retrieval of land surface emissivities over partially vegetated surfaces from satellite data using radiative transfer models," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5003821, doi: 10.1109/TGRS.2022.3224639.

[16] C. Coll, Z. Wan, and J. M. Galve, "Temperature-based and radiance-based validations of the V5 MODIS Land Surface Temperature product," *J. Geophys. Res., Atmos.*, vol. 114, no. 20, pp. 1–15, Oct. 2009, doi: 10.1029/2009jd012038.

[17] Z. Wan and J. Dozier, "A generalized split-window algorithm for retrieving land-surface temperature from space," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 892–905, Jul. 1996, doi: 10.1109/36.508406.

[18] G. C. Hulley and S. J. Hook, "Generating consistent Land Surface Temperature and emissivity products between ASTER and MODIS data for Earth science research," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1304–1315, Apr. 2011, doi: 10.1109/TGRS.2010.2063034.

[19] G. Hulley, N. Malakar, T. Hughes, T. Islam, and S. Hook, "Moderate resolution imaging spectroradiometer (MODIS) MOD21 Land Surface Temperature and emissivity algorithm theoretical basis document (ATBD)," JPL, Pasadena, CA, USA, Tech. Rep. JPL Publication 12-17, Mar. 2016.

[20] A. Gillespie, S. Rokugawa, T. Matsunaga, J. S. Cothern, S. Hook, and A. B. Kahle, "A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1113–1126, Jul. 1998, doi: 10.1109/36.700995.

[21] A. Berk, G. P. Anderson, P. K. Acharya, and E. P. Shettle, "MODTRAN5. 2.0. 0 user's manual," Air Force Res. Lab., Spectral Sci., Burlington, MA, USA, 2008.

[22] E. Valor and V. Caselles, "Mapping land surface emissivity from NDVI: Application to European, African, and South American areas," *Remote Sens. Environ.*, vol. 57, no. 3, pp. 167–184, Sep. 1996, doi: 10.1016/0034-4257(96)00039-9.

[23] X. Wu and W. L. Smith, "Emissivity of rough sea surface for 8–13 μm: Modeling and verification," *Appl. Opt.*, vol. 36, no. 12, p. 2609, Apr. 1997, doi: 10.1364/ao.36.002609.

[24] R. Niclòs, V. Caselles, E. Valor, C. Coll, and J. M. Sánchez, "A simple equation for determining sea surface emissivity in the 3–15 μm region," *Int. J. Remote Sens.*, vol. 30, no. 6, pp. 1603–1619, Mar. 2009, doi: 10.1080/01431160802541523.

[25] M. Legrand, C. Pietras, G. Brogniez, M. Haeffelin, N. K. Abuhassan, and M. Sicard, "A high-accuracy multiwavelength radiometer for in situ measurements in the thermal infrared. Part I: Characterization of the instrument," *J. Atmos. Ocean. Technol.*, vol. 17, no. 9, pp. 1203–1214, Sep. 2000.

[26] R. Niclos et al., "Evaluation of Landsat-8 TIRS data recalibrations and Land Surface Temperature split-window algorithms over a homogeneous crop area with different phenological land covers," *ISPRS J. Photogramm. Remote Sens.*, vol. 174, pp. 237–253, Apr. 2021, doi: 10.1016/j.isprsjprs.2021.02.005.

[27] R. Niclòs, J. M. Galve, J. A. Valiente, M. J. Estrela, and C. Coll, "Accuracy assessment of Land Surface Temperature retrievals from MSG2-SEVIRI data," *Remote Sens. Environ.*, vol. 115, no. 8, pp. 2126–2140, Aug. 2011, doi: 10.1016/j.rse.2011.04.017.

[28] G. Sepulcre-Cantó, P. J. Zarco-Tejada, J. C. Jiménez-Muñoz, J. A. Sobrino, E. D. Miguel, and F. J. Villalobos, "Detection of water stress in an olive orchard with thermal remote sensing imagery," *Agricult. Forest Meteorol.*, vol. 136, nos. 1–2, pp. 31–44, Jan. 2006, doi: 10.1016/j.agrformet.2006.01.008.

[29] S. B. Idso, R. D. Jackson, W. L. Ehrler, and S. T. Mitchell, "A method for determination of infrared emittance of leaves," *Ecology*, vol. 50, no. 5, pp. 899–902, Sep. 1969, doi: 10.2307/1933705.

[30] C. Coll, V. García-Santos, R. Niclòs, and V. Caselles, "Test of the MODIS Land Surface Temperature and emissivity separation algorithm with ground measurements over a rice paddy," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3061–3069, May 2016, doi: 10.1109/TGRS.2015.2510426.

[31] R. Niclòs, L. Pérez-Planells, C. Coll, J. A. Valiente, and E. Valor, "Evaluation of the S-NPP VIIRS Land Surface Temperature product using ground data acquired by an autonomous system at a rice paddy," *ISPRS J. Photogramm. Remote Sens.*, vol. 135, pp. 1–12, Jan. 2018, doi: 10.1016/j.isprsjprs.2017.10.017.

[32] R. Niclòs, M. Perelló, J. Puchades, C. Coll, and E. Valor, "Evaluating Landsat-9 TIRS-2 calibrations and Land Surface Temperature retrievals against ground measurements using multi-instrument spatial and temporal sampling along transects," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 125, Dec. 2023, Art. no. 103576, doi: 10.1016/j.jag.2023.103576.

[33] L. Pérez-Planells et al., "Validation of Sentinel-3 SLSTR Land Surface Temperature retrieved by the operational product and comparison with explicitly emissivity-dependent algorithms," *Remote Sens.*, vol. 13, no. 11, p. 2228, Jun. 2021, doi: 10.3390/rs13112228.

[34] H. Li et al., "Temperature-based and radiance-based validation of the collection 6 MYD11 and MYD21 Land Surface Temperature products over barren surfaces in northwestern China," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1794–1807, Feb. 2021, doi: 10.1109/TGRS.2020.2998945.

[35] Z. Wan, "New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product," *Remote Sens. Environ.*, vol. 140, pp. 36–45, Jan. 2014, doi: 10.1016/j.rse.2013.08.027.

[36] S.-B. Duan, Z.-L. Li, H. Wu, P. Leng, M. Gao, and C. Wang, "Radiance-based validation of Land Surface Temperature products derived from collection 6 MODIS thermal infrared data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 70, pp. 84–92, Aug. 2018, doi: 10.1016/j.jag.2018.04.006.

[37] H. Li et al., "Evaluation of the VIIRS and MODIS LST products in an arid area of Northwest China," *Remote Sens. Environ.*, vol. 142, pp. 111–121, Feb. 2014, doi: 10.1016/j.rse.2013.11.014.

[38] S.-B. Duan et al., "Validation of Collection 6 MODIS Land Surface Temperature product using in situ measurements," *Remote Sens. Environ.*, vol. 225, pp. 16–29, May 2019.

César Coll received the B.Sc., M.Sc., and Ph.D. degrees in physics from the University of Valencia, Valencia, Spain, in 1988, 1992, and 1994, respectively.

He has been a Full Professor of earth physics with the Department of Earth Physics and Thermodynamics, Faculty of Physics, University of Valencia, since 2010, where he is currently the Director of the remote sensing master's degree. He has published more than 70 scientific articles in international journals and 100 conference papers. His research interest focuses on the physical processes and applications of thermal-infrared (TIR) remote sensing and ground validation of land surface temperature (LST) and emissivity.

Raquel Niclòs received the B.Sc., M.Sc., and Ph.D. degrees in physics from the University of Valencia (UV), Valencia, Spain, in 2000, 2002, and 2005, respectively.

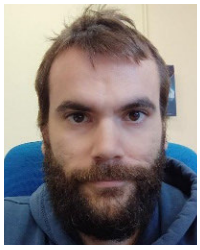
She has more than 20 years of research experience in Earth physics. She is currently an Associate Professor with the Department of Earth Physics and Thermodynamics, UV, since 2017. She has been a Supervisor of seven Ph.D. theses, the Principal Investigator of four research and development and innovation (R&D&i) projects in competitive calls and five contracts with outstanding companies, and a member of five international scientific committees. She has published more than 45 articles in journals including the Journal Citation Reports (JCR), with 90% (50%) in the top tercile (top decile) positions. She has also published more than 100 works as non-JCR papers, book chapters, and conference proceedings. She was an Editor of an international book about thermal remote sensing advances and holds a patent. Her research interests mainly focus on thermal-infrared (TIR) remote sensing, atmospheric and emissivity corrections, calibration of TIR sensors, and validation of satellite products of the land and sea surface temperatures, but also other related climate essential variables, meteorological and climate applications, and trends and impacts associated with climate change.

Jesús Puchades received the B.Sc. degree in electronic telecommunications engineering and the M.Sc. degree in remote sensing from the University of Valencia, Valencia, Spain, in 2015 and 2020, respectively, where he is currently pursuing the Ph.D. degree in remote sensing.

As a member of the Thermal Remote Sensing Group, University of Valencia, he was involved in several Spanish research projects and land surface temperature (LST) validation campaigns. His research interests include the retrieval and validation of land surface temperature/emissivity from satellite data.

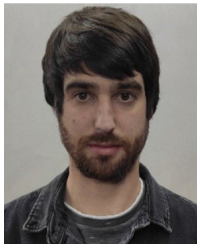
Vicente García-Santos received the B.Sc., M.Sc., and Ph.D. degrees in physics from the University of Valencia, Valencia, Spain, in 2007, 2010, and 2013, respectively.

He is currently an Assistant Lecturer at the University of Valencia. He has authored more than 27 articles published in international journals and more than 50 conference papers. His research interests include the physical processes of thermal infrared remote sensing with special attention to the retrieval and validation of land surface temperature and emissivity products, characterization and correction of atmospheric effects, and estimation of energy fluxes exchange between surface and atmosphere.



Martín Perelló received the B.Sc., M.Sc., and Ph.D. degrees in physics from the University of Valencia (UV), Valencia, Spain, in 2014, 2015, and 2020, respectively.

He is a member of the Thermal Remote Sensing Group, University of Valencia, where he has participated in international and Spanish land surface temperature (LST) validation campaigns. His research interests include the calibration and validation of LST from satellite data and the study of LST trends associated with climate change.



Lluís Pérez-Planells received the B.Sc. degree in physics, the M.Sc. degree in remote sensing, and the Ph.D. degree in remote sensing from the University of Valencia, Valencia, Spain, in 2014, 2015, and 2021, respectively.

He is a member of the Thermal Remote Sensing Group, University of Valencia. He is currently a Post-Doctoral Researcher with Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, where he is the Main Validation Scientist for the LST_cci project of the European Space Agency (ESA), responsible for validating and intercomparing a broad range of land surface temperature (LST) satellite products generated within the project. He is also a Team Member of the Copernicus LAW project, for which he developed software for continuously analyzing and monitoring in situ LST from five new LST validation stations, i.e., by evaluating the associated uncertainties and the calibration and stability of the deployed field radiometers. He was involved in several Spanish research projects and land surface temperature (LST) validation campaigns. His research focused on the modeling of thermal-infrared (TIR) radiative transfer and the emissivity of vegetated surfaces.