International Journal of Applied Earth Observation and Geoinformation 125 (2023) 103576

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Evaluating Landsat-9 TIRS-2 calibrations and land surface temperature retrievals against ground measurements using multi-instrument spatial and temporal sampling along transects

Raquel Niclòs *, Martín Perelló, Jesús Puchades, César Coll, Enric Valor

Thermal Remote Sensing Group, Department of Earth Physics and Thermodynamics, Faculty of Physics, University of Valencia, 50, Dr. Moliner, E-46100 Burjassot, Spain

A R T I C L E I N F O

Keywords:
Landsat 9 TIRS-2
Calibration
Land surface temperature
Validation
Ground TIR data
Multi-instrument spatio-temporal sampling

A B S T R A C T

High-resolution TIR data are essential for a wide range of studies, e.g., in environmental monitoring, urban effects, water stress, agricultural productivity, and land/water resources. Landsat series provides high-resolution TIR data, but data accuracy must be ensured to contribute to these fields. After the calibration problems of Landsat 8 (L8) TIRS due to a stray light effect, efforts were made in the design of Landsat 9 (L9) TIRS-2 to mitigate this effect. In any case, field calibration and validation activities are necessary to evaluate the accuracy of recently launched sensors. These activities require high-quality ground TIR radiance data that are representative at satellite spatial resolution and that quantify spatial–temporal fluctuations at the experimental sites concurrently to satellite overpasses. However, these requirements are not always accomplished in the literature.

This paper provides vicarious calibrations of the L9 TIRS-2 data with accurate ground measurements acquired using multi-instrument spatial and temporal sampling in a rice paddy area from December 2021 to August 2023, in order to contribute to the calibration and validation of L9 TIRS-data.

In March 1, 2023 L9 TIRS-2 scenes started to be reprocessed and calibration was updated. Both the original calibration and that after reprocessing were evaluated. Close results were obtained for band 10 (b10). When comparing satellite brightness temperatures and those determined from in situ LSTs, systematic differences of −0.3 K were shown, with root-mean-square differences (RMSDs) of 0.5 K. Results for band 11 showed a bias of −0.1 K for the original data and of 0.2 K after the reprocessing, with RMSDs of 0.6 K.

The L2 LST product generated operationally from b10 data was also evaluated, together with alternative single-channel (SC) algorithms. A bias (RMSD) of 0.4 K (1.1 K) was obtained for the product. However, negative biases were shown both with the radiative transfer equation used on b10 radiances and the alternative SC algorithm.

Finally, we validated six types of split-window (SW) algorithms applied to L9 TIRS-2 data before and after the reprocessing. A systematic difference of about −0.6 K between SW LST biases when using the reprocessed and original data was observed. In any case, for most of the algorithms with original and reprocessed data, LST biases and standard deviations were within the recommended threshold of 1 K, which proves the good performance of L9 TIRS-2 data to monitor LST at high spatial resolution.

1. Introduction

LANDSAT 9 (L9) has been in orbit since late September 2021, and operational since February 2022. Nowadays L9 together with Landsat 8 (L8) provide a full coverage every 8 days. Thermal Infrared Sensor 2 (TIRS-2) is onboard L9 together to the Operational Land Imager 2 (OLI-2). It has two thermal-infrared (TIR) bands like the TIRS in L8. A primary radiometric calibration change was introduced with the addition of baffling in the L9 TIRS-2 telescope to minimize the stray light effect found in the case of L8 TIRS (Barsi et al., 2022; Montanaro et al., 2022). On-orbit calibrations have been performed, and according to Barsi et al. (2022), TIRS-2 has shown excellent radiometric performance, with noise level below 0.07 K for both bands. Additionally, lunar scans and vicarious calibration data showed that the stray light effect has been mitigated (Barsi et al., 2022; Montanaro et al., 2022). Up to March 1, 2023, the L9 TIRS-2 absolute calibration come from the prelaunch one, with a

* Corresponding author.
E-mail address: raquel.niclos@uv.es (R. Niclòs).

https://doi.org/10.1016/j.jag.2023.103576
Received 29 August 2023; Received in revised form 2 November 2023; Accepted 18 November 2023
Available online 28 November 2023

small response change introduced after the March 2022 cryocooler electronics reset (Barsi et al., 2022). However, L9 collection 2 (C2) data started to be reprocessed on March 1, 2023 based on the first calibration results obtained by the USGS/NASA Calibration and Validation team using ground data acquired in inland waters. According to them, the updates after TIRS-2 data reprocessing were: (i) updates to the absolute radiances, with the largest adjustment of around 0.7 K in band 11 (b11). (ii) Adjustments to improve TIRS-2 to OLI-2 co-instrument alignment. (iii) 16 detectors in TIRS-2 band 10 (b10) and 3 detectors in b11 were replaced with on-board redundant detectors to reduce striping due to detector response changes (Barsi et al., 2022).

L9 TIRS-2 can provide valuable land surface temperature (LST) records at high spatial resolution (100 m, resampled to 30 m). LST is a key variable to understand land–atmosphere physical processes, and thus for climatological and hydrological applications (Anderson et al., 2012; Good et al., 2022; Li et al., 2023). It was defined as an essential climate variable by the Global Climate Observing System (GCOS) (GCOS, 2016), and also by space agencies such as NASA and ESA, the latter through the Climate Change Initiative. Additionally, high-resolution LST retrievals are valuable in the study of: natural hazards, thermal anomalies, fire detection and severity, urban heat island effects, crop irrigation needs, and natural resources (Darge et al., 2019; García-Llamas et al., 2019; Li et al., 2023; Peres et al., 2018; Roy et al., 2015). All these applications are the objectives of the large number of dedicated satellite missions approved for the near future, i.e., TRISHNA, SBG-TIR, LSTM, and Landsat Next (Vidal et al., 2022).

Suitable calibration and validation (CAL/VAL) activities based on independent ground-data (T-based method) are required to evaluate the accuracy of both the TIRS-2 calibration and the Level 2 (L2) single-channel (SC) derived LST product, and also the suitability of split-window (SW) algorithms to be used on TIRS-2 data. Significant systematic and/or random uncertainties in the calibration of the TIRS-2 bands and in the LSTs retrieved from TIRS-2 radiance data would prevent their used for the mentioned applications. This was the case of L8 TIRS band 11 data before the systematic application of the stray light correction method proposed by Montanaro et al. (2015) in view of the CAL/VAL results. The Landsat calibration team advised against its use in SC and SW algorithms to retrieve LSTs.

First calibration evaluations for L9 TIRS-2 were performed over water surfaces in Barsi et al. (2022) between November 2021 and April 2022, which indicated possible small underestimations of less than 0.2 K in b10 and 0.7 K in b11. They used single-band TIR radiometers set at buoys in Lake Tahoe and Salton Sea, but also mined data from the NOAA National Data Buoy Center. The results were used for the reprocessing starting on March 1, 2023. However, further CAL/VAL experiments carried out over land surfaces, using ground data acquired at homogeneous sites, would allow more reliable evaluation results to be obtained. Meng et al. (2022) and Ye et al. (2022) evaluated L9 TIRS-2 LSTs retrieved before reprocessing using a dataset of temperatures derived from pyrgeometer data as reference. Meng et al. (2022) indicated root-mean-square errors (RMSEs) $\geq$ 3 K in L2 LSTs, even after filtering outliers. Ye et al. (2022) showed a RMSE of 1.8 K for L2 LSTs. However, previous studies indicated that pyrgeometer data were not accurate enough for CAL/VAL evaluations, since: (i) only the instrument uncertainty lead to a 1 K uncertainty in LST, (ii) substantial differences were shown when comparing LSTs determined from data measured by TIR radiometers and by pyrgeometers, and (iii) residual errors were consistently observed when comparing satellite-retrieved LSTs and those obtained from pyrgeometer data (Gerace et al., 2020; Guillevic et al., 2012; Krishnan et al., 2020; Niclòs et al., 2021). Thus, more robust ground-measured datasets were recommended for validations (Guo et al., 2020; Niclòs et al., 2021). Ground datasets should be acquired over thermally-homogeneous sites with TIR radiometers (Hulley et al., 2021), and possible spatial and temporal fluctuations should be evaluated during satellite overpasses (Lagouarde et al., 2019).

The objective of the paper is to provide CAL/VAL results for the L9 TIRS-2 data using accurate ground TIR data acquired with calibrated TIR radiometers carried on along transects in a rice paddy area, known as Valencia test site (Coll et al., 2012, 2005; Niclòs et al., 2021, 2018). The land surface at the site changes across the year, which is an advantage for these activities since allow us to measure in a wide range of temperatures and conditions. This site was formerly used to evaluate the performances of several satellite sensors (Coll et al., 2012, 2005; Niclòs et al., 2021, 2018), e.g., L8 TIRS bands when the stray light effect was a problem and after the correction method (Niclòs et al., 2021).

The contributions and novelties of the present paper include: (1) Vicarious calibrations of the new L9 TIRS-2 over homogeneous land covers, both for the original calibrated data (hereafter OC data; with absolute calibration based on the prelaunch calibration) and the March 2023 reprocessed data (hereafter RP data; with the calibration adjustment). (2) Use of accurate ground TIR data acquired with multi-instrument spatial and temporal sampling to quantify the site thermal spatial variability and possible high-frequency temporal fluctuations induced by atmospheric turbulence during each L9 overpass. (3) Validation of the operational L2 product, together to alternative SC methods, both for OC and RP data. (4) Evaluation of six types of SW algorithms applied to L9 TIRS-2 data in order to show the change in uncertainties introduced by the reprocessing, and test whether the accuracies of the algorithms are within recommended thresholds for climate studies.

The next section describes the experimental site, the calibration of the TIR radiometers used to acquire measurements and the procedure followed to derive LSTs from the radiance data measured. Section 3 describes the methodology followed to evaluate the calibrations of TIRS-2 b10 and b11 and the derived L2 LST product, including atmospheric radiative transfer simulations and emissivities. This section also includes a list of SW algorithms additionally validated. The calibration results both for OC and RP data are shown and discussed in Section 4.1. The validation results of the L2 LST product are shown in Section 4.2, and the results when using SW algorithms are analyzed in Section 4.3. Lastly, we summarize the most important findings in Section 5.

2. Ground data

2.1. Test site

Ground TIR radiance data were acquired in the Valencia Test Site concurrently to cloud-free L9 overpasses from December 2021 to August 2023. The site is a rice paddy area with several land surfaces as a consequence of the rice crop phenology, varying from flooded surface and wet and dry bare soils to full vegetation cover, as described in Niclòs et al. (2021). The site is covered by WRS-2 paths 198 and 199 of row 33. Fig. 1 shows L9 OLI-2 scenes at the site on May 26, 2022 (ploughed dry bare soil), August 14, 2022 (full vegetation), and January 21, 2023 (flooded soil). Fig. 2 shows photographs for the three surface conditions at the site taken in measurement days. All land covers were shown to be thermally homogeneous at various spatial scales (Coll et al., 2012, 2005; Niclòs et al., 2018). LST standard deviations (SDs) obtained from radiometer measurements along 200 m transects in the site were generally lower than 0.5 K, and up to 0.9 K in some cases for bare soils (Niclòs et al., 2021, 2018). Therefore, the site is interesting for CAL/VAL purposes under three very different but appropriate surface conditions, which permit to cover a wide range of temperatures throughout the year.

2.2. Ground TIR radiance measurements

Four Cimel Electronique TIR multichannel radiometers, model CE-312 (Brogniez et al., 2003), were used to acquire TIR radiance data. These radiometers were hand-held, distributed over the area and carried on along transects, covering the maximum possible extension within 5 min around the L9 overpass time. Each radiometer took measurements each 5–10 m, with about 25–30 measurements within the 5 min per

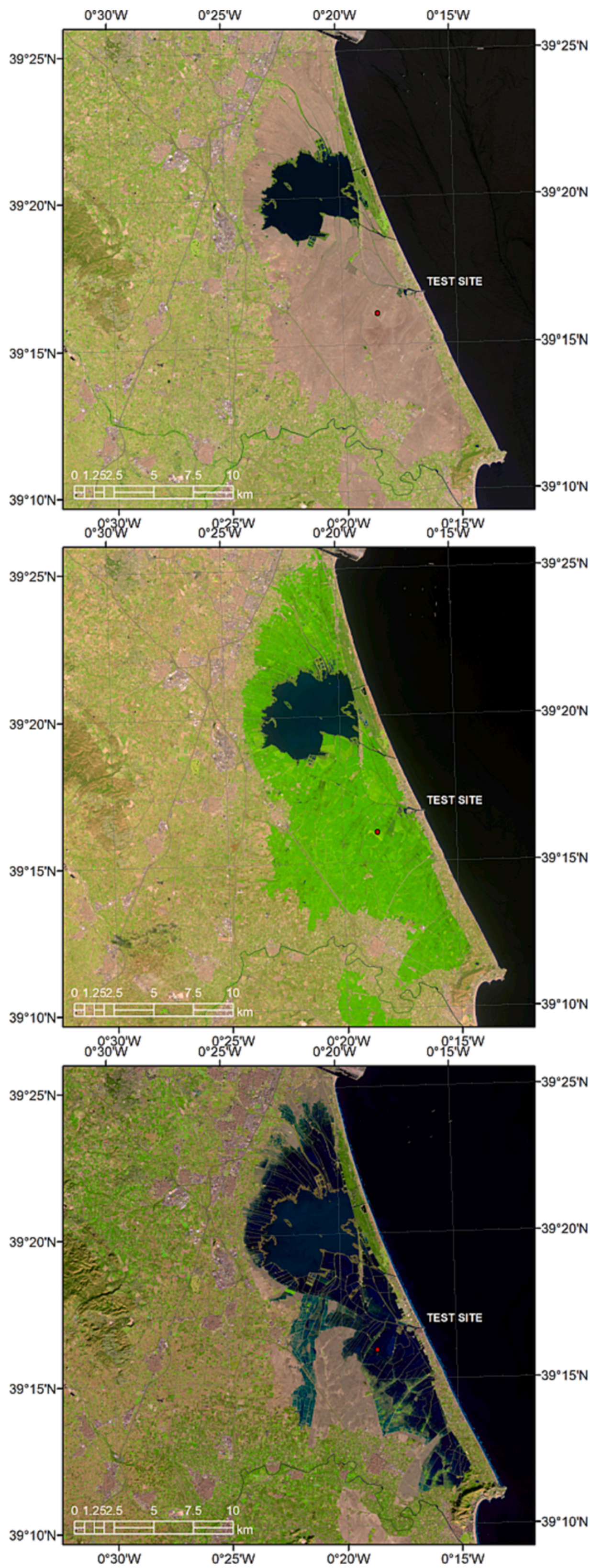


Fig. 1. L9 OLI-2 color composites (RGB 654) at the site for three of the measurement days: May 26, 2022 (ploughed dry bare soil), August 14, 2022 (full vegetation), and January 21, 2023 (flooded soil).

radiometer. Thus, using 4 radiometers, the total number of measurements taken was 100–120 per L9 overpass. This sampling was done to assess the thermal variability, which is one of the main sources of uncertainty for ground-measured LSTs. According to Lagouarde et al. (2019), the spatial thermal variability of land surfaces and temporal fluctuations induced by atmospheric turbulence can only be assessed whether measurements of several radiometers, acquiring data sufficiently spaced apart to avoid correlations, are averaged over the measurement time. Thus, such measurement protocol applied here increases the accuracy of the evaluation results.

Ground measurements were done following this procedure concurrently with the acquisition of 23 L9 scenes with cloud-free conditions over the site within the abovementioned period.

The ground radiometers acquired surface-leaving radiance in each band i as follows:

$$L_{surf,i} = \varepsilon_i B_i(LST) + (1-\varepsilon_i) L_{ia,hem}\downarrow \tag{1}$$

where B_i is the band-averaged Planck's radiance function, ε_i is emissivity, and $L_{i\,a,hem}^{\downarrow}$ is the atmospheric downwelling irradiance divided by π in the case of Lambertian surfaces (i.e., bare soil and full vegetation). $L_{i\,a,hem}^{\downarrow}$ was measured at the site using a Labsphere Infragold Reflectance Target IRT-94–100 with a high reflectivity in the TIR (Niclòs et al., 2021). However, specular reflection was considered for flooded surface, and $L_{i\,a,hem}^{\downarrow}$ was replaced by $L_{i\,a}^{\downarrow}$ in Eq. (1), which was measured directly at zenith. The relative percentage difference between $L_{i\,a,hem}^{\downarrow}$ and $L_{i\,a}^{\downarrow}$ (when divided by the former) is from 36 % to 42 % at 10.9 μm for the days we measured flooded soils, with a ratio between them from 1.5 to 1.7.

2.3. Calibration of ground radiometers

The six TIR bands of the CE-312 radiometers were calibrated once per year, and before and after experimental campaigns, using a Landcal P80P calibration source. In June 2022, our radiometer calibrations were also evaluated in an international TIR radiometer inter-comparison funded by ESA within the FRM4STS project (Yamada et al., 2023a, 2023b). During the inter-comparison experiment, TIR radiometers usually used for validation purposes were compared with reference blackbodies (BB), with uncertainties of 0.04–0.13 K for the measurement range (Yamada et al., 2023a, 2023b). They were set at nominal temperatures from −30 to 50 °C. The uncertainty budget of our radiometers was assessed as described in Coll et al. (2019) for the previous experiment in 2016, and combined uncertainties (i.e., total root sum squares) lower than 0.2 K were obtained from the 2022 calibration data. Calibration equations were also estimated from the experiment results for each radiometer and band. Since the relationship between radiance differences and radiometer digital number differences is linear, we used linear recalibration equations between BB minus radiometer-measured brightness temperatures ($T_{i,BB}$-$T_{i,m}$) and radiometer-measured minus internal detector temperature ($T_{i,m}$-T_d) (Coll et al., 2019). $T_{i,m}$ were measured with each band calibration provided originally by the manufacturer. Fig. 3 plots the results for one of the radiometers and band as an example, the solid line being the regression equation line.

2.4. Ground emissivity measurements

Unlike some previous studies, ε_i were neither estimated from emissivity datasets (e.g., ASTER-GED) or spectral libraries (e.g., ECOSTRESS) nor from threshold methods in terms of spectral vegetation indexes (Chatterjee et al., 2017; Gerace et al., 2020; Guo et al., 2020; Meng et al., 2019; Rozenstein et al., 2014; Wang et al., 2023; Yu et al., 2014). When possible, surface emissivity data acquired concurrently with a satellite sensor overpass is the best option (Rozenstein et al., 2014), as emissivity depends on the specific element types at the site and also varies with seasons and modifications in land use and soil moisture (Mira et al.,

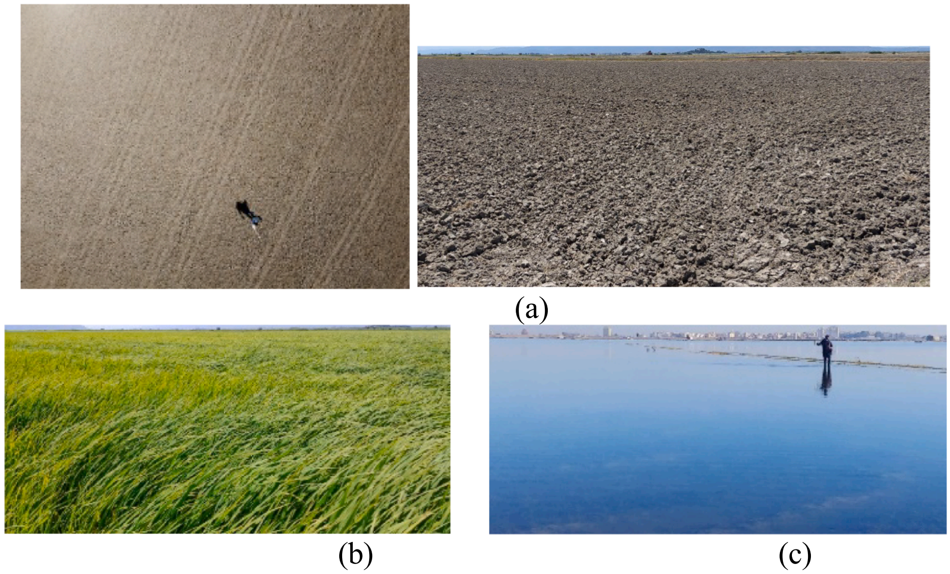


Fig. 2. Photographs taken in the experimental area with the three totally disparate surface conditions: (a) bare soil, (b) complete rice coverage, and (c) flooded soil.

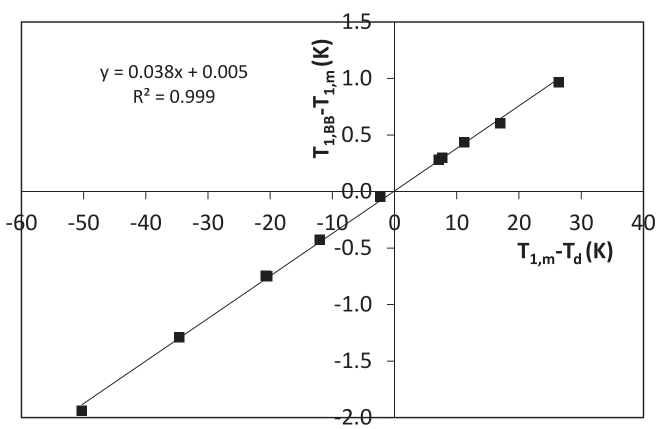


Fig. 3. Example of recalibration for band 1 of one of the radiometers used.

2010). Therefore, ε_i were measured at the Valencia test site to correct the ground-measured TIR radiances. Different measurement methods were used. The Temperature-Emissivity Separation method (Hulley and Hook, 2009) was used for wet to dry bare soils, since it works more accurately due to the higher emissivity spectral variations of this kind of surfaces. With vegetation, ε_i were measured with the Box Method since the emissivity spectrum was flatter (Niclòs et al., 2021, 2018). For water surfaces, ε_i values were obtained as described in Niclòs et al. (2014). Average ε_i values of 0.99 and 0.985 were measured (at 10.9 μm) for flooded soil and full vegetation, respectively, at the site. However, values of 0.972 were obtained for dry bare soil. ε_i uncertainties from 0.004 to 0.006 were determined for the two first types of covers, and from 0.004 to 0.011 for bare soils.

2.5. Ground LSTs and uncertainties

Eq. (1) was inverted in order to obtain the LSTs from the ground-measured radiances and corresponding measurements of surface emissivities and atmospheric downwelling radiances. Finally, we calculated average values of the LSTs retrieved from the TIR radiances measured by the different CE-312 radiometers for each L9 overpass. The average ground LSTs ranged from 279 K to 314 K for the whole study period. Additionally, the uncertainty budget was assessed for the LST retrieval, and the different uncertainty sources were added together in quadrature (Niclòs et al., 2021). An uncertainty lower than 0.2 K was estimated for the calibration of the radiometer bands used. A spatial and temporal variability usually from 0.3 to 0.5 K was determined from the LSTs measured by the different radiometers along transects within 5 min of each L9 overpass. The largest uncertainty propagated through Eq. (1) was the emissivity one, of around 0.4–0.6 K in LSTs. Finally, ground LST total uncertainties of 0.6–1.0 K were determined for the study period.

The same dataset of ground-measured LSTs was used for evaluating the calibration of L9 TIR bands (Section 3.1) and for the validation of SC and SW LST products (Sections 3.2 and 3.3).

3. Methodology

3.1. Evaluation of the L9 TIRS-2 calibrations

The vicarious calibration involved the evaluation of L9 TIRS-2 data against those determined from ground data. TOA radiances, L_i, were simulated as follows. First, we calculated surface radiances from ground-measured LSTs using Eq. (1) with the band-integrated magnitudes for TIRS-2 bands 10 and 11. Second, simulated L9 TOA radiances were obtained from the surface radiances with the radiative transfer equation (RTE):

$$L_i = \tau_i L_{surf,i} + L_{ia}\uparrow \tag{2}$$

where τ_i and $L_{i\,a}^{\uparrow}$ are the transmittance and upwelling radiance for the whole atmosphere, respectively. They, together with $L_{i\,a,hem}^{\downarrow}$, were estimated for the atmospheric condition of each L9 overpass and then integrated for TIRS-2 bands. For that, we used MODTRAN (Berk et al., 2008) simulations with NCEP atmospheric profiles interpolated for the site and overpass time. We extracted the four NCEP grid corners surrounding the site location for the two times before and after the overpass time. First, the corner profiles were spatially interpolated to obtain the profiles at the site location for each time, and then the resulting profiles were temporally interpolated for the acquisition time. The interpolated profiles were then completed from their maximum heights (of around 30 km above sea level) to 100 km from the MODTRAN standard atmospheres most appropriate for the site location and L9 overpass time (i.e., midlatitude summer or winter profiles). An inverse distance weighted interpolation with $1/d_j^2$ as weighting function was used for the spatial interpolation, d_j being the distance from the site location to the NCEP grid corners (Coll et al., 2012). Additionally, ε_i values required in Eq. (1)

were those measured in the experimental area (Section 2). The calibration was evaluated in terms of L_i, but also brightness temperatures, T_i. A flowchart of the vicarious calibration procedure is shown in Fig. 4a.

L9 TIRS-2 scenes (C2) were downloaded, both before and after the March 2023 reprocessing (i.e., OC and RP data), using the tool provided in earthexplorer.usgs.gov. We used the raster tool of the Sentinel Application Platform (SNAP) to extract pixel values of these scenes over the experimental site. The pixel values were then analyzed using a Matlab code developed to assess band 10 and 11 radiances by applying the radiometric calibration to convert raw digital numbers to radiances with scale and offset values provided in the metadata of the scenes. Brightness temperatures were calculated from radiances using the Planck's function with band-dependent coefficients. Then, evaluations against ground data were done using the single value of the pixel covering the site. We also calculated average values and SDs for a pixel array covering 5 × 5 100-m pixels (~17 × 17 30-m pixels) (Niclòs et al., 2021). This array size implies a good compromise between a substantial number of 30 m pixels to compute SDs and keeping a homogeneous footprint for the array. The rice paddy parcels have usually distances between narrow adjacent paths or irrigation ditches from 500 m to 750 m. Thus, 5 x 5 100-m pixels are similar size, and the array footprint is not covering these paths and ditches. In any case, we also tested the use of averages and SDs for 3x3 100-m pixels to compare results.

After the L_i and T_i calibration, we also used the RTE (with Eqs. (1) and (2)) to assess LST from b10 and b11 L_i, separately, and then they were evaluated directly against ground LSTs.

Four uncertainty sources were considered to determine the uncertainties in the LSTs retrieved in this way (Niclòs et al., 2021): (i) those of measured ε_i (from 0.004 to 0.011), (ii) uncertainties of the atmospheric variables both due to those in NCEP profiles and MODTRAN simulations (0.4 K and 0.8 K for TIRS-2 b10 and b11) (Coll et al., 2012), (iii) uncertainties in T_i, as $NE\Delta T/n.pixels^{1/2}$, where n.pixels is the number of pixels in the array (Ghent et al., 2019), and (iv) the array variability, as the SD of the pixels considered. $NE\Delta T < 0.1$ K was shown for both TIRS-2 bands (Barsi et al., 2022; Pearlman et al., 2022). These four uncertainty contributions led to LST uncertainties from 0.5 to 0.8 K for b10 and from 0.8 to 1.0 K for b11. These uncertainty sources were also considered to determine uncertainties in L_i and T_i assessed from in situ LSTs using Eqs. (1) and (2) directly, thus obtaining T_{10} uncertainties from 0.7 to 1.0 K and T_{11} uncertainties from 0.9 to 1.1 K.

3.2. L2 LST product validation

The L9 L2 LST product available in C2 makes use of the RTE, i.e., uses a SC technique on TIRS-2 b10 radiances. It uses ε_{10} values estimated from the ASTER-GED and MODTRAN simulations of the atmospheric parameters (τ_{10}, $L^{\uparrow}_{10\ a}$, $L^{\downarrow}_{10\ a,hem}$) as input data (Malakar et al., 2018; Meng et al., 2022). The latter were performed using the GEOS-5 FP-IT atmospheric data (https://gmao.gsfc.nasa.gov/GMAO_products/NRT_products.php). ASTER-GED emissivities, i.e., average emissivities generated from cloud-free ASTER images between 2000 and 2008 at ~ 100 m (3 arc sec) resolution, are temporally and spectrally adjusted for TIRS-2 data. Temporal adjustments are based on changes in vegetation and snow covers by using related spectral indexes (Malakar et al., 2018). L9 scenes of the product were also downloaded before and after the March 2023 reprocessing, i.e., for OC and RP data.

Additionally, an alternative SC method suggested by Jiménez-Muñoz et al. (2014) was used for comparison. We applied this method to at-sensor L_i and T_i, data in b10 but also in b11 (hereafter referred as SC_JM_10 and SC_JM_11, respectively). Additional input values were ε_i measured at the site and the atmospheric total water vapor content, W, estimated for each scene from interpolated NCEP profiles (Fig. 4a). The method provides second degree polynomial regression equations between atmospheric variables and W, which facilitate the algorithm application for potential users (i.e., without needing atmospheric simulations of τ_i, $L^{\uparrow}_{i\ a}$, $L^{\downarrow}_{i\ a,hem}$). Coefficients for these equations were only given for b10 in Jiménez-Muñoz et al. (2014), but Guo et al. (2020) provided coefficients for b11. We used both of them. However, this approximation can involve higher uncertainties. Yu et al. (2014) compared the L8 results of this SC algorithm with those when using directly the RTE, and concluded that LSTs estimated from direct RTE calculations were more accurate. Jiménez-Muñoz et al. (2014) evaluated the input W uncertainty and the impact in the LST uncertainty with the algorithm, of up to 1 K. This input variable together with the emissivity one and the algorithm regression uncertainty lead to considerable LST uncertainties compared to those of the RTE.

3.3. Evaluation of SW algorithms

As summarized in Niclòs et al. (2021), five types of SW algorithms were fit for L8, each of them with several versions due to regression coefficients obtained from different simulation datasets and different

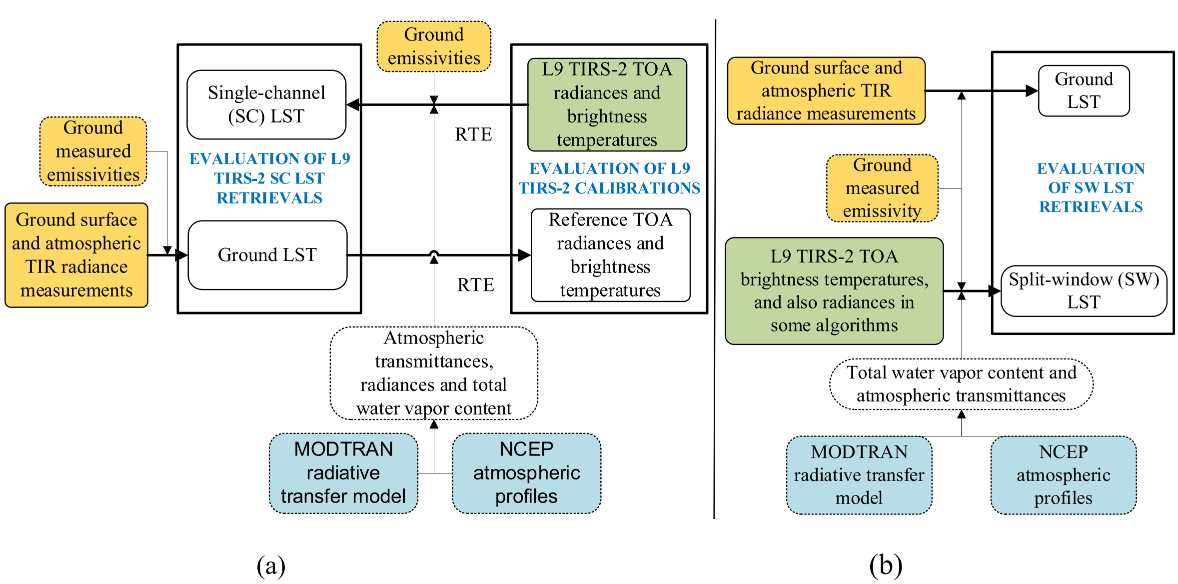


Fig. 4. Flowcharts of the procedure followed for evaluating: a) both the L9 TIRS-2 calibrations and the LSTs retrieved from L9 TIRS-2 data inverting the RTE, but also SC algorithms, and b) SW algorithms.

parametric equations. As indicated in Ye et al. (2022), b10 and b11 of TIRS-2 and TIRS are spectrally extremely close, and the identical algorithm is used to generate the L2 LST product from L9 and L8 data. Thus, we first evaluated these five SW algorithms for L9 TIRS-2.

(i) The first algorithm, proposed by Jiménez-Muñoz et al. (2014), is quadratic on brightness temperature differences, T_{10} - T_{11} (see in Table 1). We applied this algorithm both with the regression coefficients provided in Jiménez-Muñoz et al. (2014) (hereafter referred to as SW1), and also with the regression coefficients obtained in Meng et al. (2019) per subranges of W (hereafter SW1_M_W) and also for any W $\leq$ 7 cm (SW1_M).

(ii) The second class of algorithms is based on that designed by Wan and Dozier (1996) with refinements in Wan (2014), and coefficients obtained from regression by:

1) Meng et al. (2019) for any W $\leq$ 7 cm (hereafter SW2_M) and per subranges of W (SW2_M_W).

2) Du et al. (2015) for any W $\leq$ 6.3 cm (SW2_D) and per subranges of W (SW2_D_W), and also refined by Guo et al. (2020) using different coefficients depending on W and temperature (SW2_D_W_T).

3) Gerace et al. (2020) (SW2_G) for any W and temperature. This algorithm was proposed to generate operationally a new and alternative LST product in the next Landsat collection 3 (C3).

(iii) The third algorithm is the NOAA JPSS Enterprise SW equation with coefficients regressed per W subranges (hereafter SW3_M_W) and also for W $\leq$ 7.0 cm (SW3_M) by Meng et al. (2019).

(iv) The fourth algorithm was proposed by Yu et al. (2014) for TIRS (hereafter SW4) similarly to that in Qin (2001) for AVHRR, which is based on assessing parameters p_i using linear regressions with TIRS TOA T_i.

(v) The fifth algorithm is that of Rozenstein et al. (2014) (SW5) also based on that in Qin et al. (2001), but p_i parameters are unused in this case and it only considers the corresponding regression coefficients.

Both SW4 and SW5 algorithm use atmospheric transmittances, τ_i, and authors provided regressions that characterize the τ_i dependence on W for a midlatitude summer profile (hereafter SW4_MLS and SW5_MLS) and the US standard atmosphere (hereafter SW4_US and SW5_US), and they recommend the US regressions to users. The regressions are quadratic in SW4 and linear in SW5. However, we also evaluated both algorithms using the τ_i values that we simulated with MODTRAN from NCEP profiles interpolated for the L9 overpasses at the site (hereafter SW4_SIM and SW5_SIM, respectively).

(vi) Here, we also considered a recently proposed method by Wang et al. (2023) (hereafter SW6) to retrieve LSTs using L9 TIRS-2 data based on the radiance-based SW algorithm proposed for MODIS (Wang et al., 2019). The algorithm uses TIRS-2 L_i instead of T_i values. The black-body equivalent surface emittance at b10, B_{10}, is first estimated using a regression equation with L_{10} and L_{11}. Then the LST is estimated from B_{10} with the RTE. We applied the algorithm using τ_i and $\varphi_i = L^{\downarrow}_{i\ a,hem} / L^{\uparrow}_{i\ a}$ values simulated with the interpolated NCEP profiles and MODTRAN (hereafter SW6_SIM), but also those estimated from W with regression equations proposed in Wang et al. (2023) using a selection of profiles from the TIGR2000 dataset (hereafter SW6_REG).

The six SW algorithms are summarized in Table 1. These algorithms, using different coefficients or τ_i estimates, were applied to L9 TIRS-2 data (both with OC and RP data) and evaluated against ground data, through Matlab code developed. As inputs in the SW algorithms, we used the ε_i values measured at the site, and the W estimates obtained from interpolated NCEP profiles. Atmospheric parameters simulated with MODTRAN were also used for the SW4_SIM, SW5_SIM and SW6_SIM versions. Fig. 4b shows the flowchart used for the validation of the SW algorithms.

To determine the uncertainties of the different algorithms, we took the next five uncertainty sources into account: the algorithm regression uncertainty provided by the authors, the propagation in the algorithm of measured ε_i uncertainties, those of W uncertainty (0.5 cm (Niclòs et al., 2021)) and band NEΔTs (as NEΔT/n.pixels$^{1/2}$), and the SD of LSTs in the array. The total LST uncertainties varied from 0.7 K to 2.0 K for the different versions of algorithm SW1, with average uncertainties around 1.2 K. Uncertainties were from 0.4 K to 1.9 K for the different SW2 versions, with averages around 0.8–1.1 K. They varied from 0.6 K to 2.0 K for SW3 versions, with averages of 1.1 K. Uncertainties from 0.4 K to 2.1 K were determined for the SW4_SIM, SW5_SIM, SW5_US and SW5_MLS, with average uncertainties of 1.0–1.1 K (with maximum uncertainties up to nearly 3.0 K for the SW4_US and SW4_MLS, which increased the averages up to 1.5 K). Finally, they were in the range of 0.6–3.0 K for the SW6_SIM and SW6_REG versions, with averages of 1.5 K.

Table 1
Summary of the SW algorithms used, including names assigned, equations for each algorithm, and references in which they were proposed.

Name	SW algorithm	Reference
SW1	$LST = T_{10} + c_0 + c_1 (T_{10} - T_{11}) + c_2 (T_{10} - T_{11})^2 + (c_3 + c_4 W)(1 - \varepsilon) + (c_5 + c_6 W)\Delta\varepsilon$	(Jimenez-Munoz et al., 2014)
SW2	$LST = c_0 + \left(c_1 + c_2\frac{1-\varepsilon}{\varepsilon} + c_3\frac{\Delta\varepsilon}{\varepsilon^2}\right)\frac{T_{10}+T_{11}}{2} + \left(c_4 + c_5\frac{1-\varepsilon}{\varepsilon} + c_6\frac{\Delta\varepsilon}{\varepsilon^2}\right)\frac{T_{10}-T_{11}}{2} + c_7(T_{10}-T_{11})^2$	(Wan and Dozier, 1996) & Wan (2014)
SW3	$LST = c_0 + c_1 T_{10} + c_2 (T_{10} - T_{11}) + c_3 \varepsilon + c_4 \varepsilon (T_{10} - T_{11}) + c_5 \Delta\varepsilon$	(Meng et al., 2019)
SW4	$LST = T_{10} + B_1 \cdot (T_{10} - T_{11}) + B_0$ $B_0 = \frac{C_{11}\cdot(1 - A_{10} - C_{10})\cdot p_{10} - C_{10}\cdot(1 - A_{11} - C_{11})\cdot p_{11}}{C_{11}\cdot A_{10} - C_{10}\cdot A_{11}}$ $B_1 = \frac{C_{10}}{C_{11}\cdot A_{10} - C_{10}\cdot A_{11}}$ $A_{10} = \varepsilon_{10}\cdot\tau_{10}$ $A_{11} = \varepsilon_{11}\cdot\tau_{11}$ $C_{10} = (1 - \tau_{10})\cdot(1 + (1 - \varepsilon_{10})\cdot\tau_{10})$ $C_{11} = (1 - \tau_{11})\cdot(1 + (1 - \varepsilon_{11})\cdot\tau_{11})$	(Yu et al., 2014)
SW5	$LST = A_0 + A_1\cdot T_{10} - A_2\cdot T_{11}$ $A_0 = E_1\cdot b_{10} - E_2\cdot b_{11}$ $A_1 = 1 + A + E_1\cdot a_{10}$ $A_2 = A + E_2\cdot a_{11}$ $A = D_{10}/E_0$ $E_1 = D_{11}\cdot(1 - C_{10} - D_{10})/E_0$ $E_2 = D_{10}\cdot(1 - C_{11} - D_{11})/E_0$ $E_0 = D_{11}\cdot C_{10} - D_{10}\cdot C_{11}$ $C_i = \varepsilon_i\cdot\tau_i(\theta)$ $D_i = [1 - \tau_i(\theta)]\cdot[1 + (1 - \varepsilon_i)\cdot\tau_i(\theta)]$	(Rozenstein et al., 2014)
SW6	$LST = \frac{c_2\cdot\lambda_{10}^{-1}}{\ln\left(\frac{c_1\cdot\lambda_{10}^{-5}}{B_{10}(LST)} + 1\right)}$ $B_{10}(LST) = A_0\cdot L_{10} + A_1\cdot L_{11} + A_2$ $A_0 = \frac{D_{11}}{C_{10}\cdot D_{11} - C_{11}\cdot D_{10}}$ $A_1 = \frac{-D_{10}}{K\cdot(C_{10}\cdot D_{11} - C_{11}\cdot D_{10})}$ $A_2 = \frac{b\cdot D_{10}\cdot(C_{11} + D_{11})}{k\cdot(C_{10}\cdot D_{11} - C_{11}\cdot D_{10})}$ $C_{10/11} = \varepsilon_{10/11}\cdot\tau_{10/11}$ $D_{10/11} = (1 - \tau_{10/11})\cdot[(1 - \varepsilon_{10/11})\cdot\tau_{10/11}\cdot\varphi_{10/11} + 1]$ $K = \frac{c_1^2\cdot\lambda_{10}^{-4}\cdot\lambda_{11}^{-6}\cdot(c_1\cdot\lambda_{10}^{-5}\cdot L_{10}^{-1} + 1)^{(\lambda_{10}\cdot\lambda_{11}^{-1} - 1)}}{\left[(c_1\cdot\lambda_{10}^{-5}\cdot L_{10}^{-1} + 1)^{(\lambda_{10}\cdot\lambda_{11}^{-1})} - 1\right]^2\cdot L_{10}^2}$ $b = \frac{c_1\cdot\lambda_{11}^{-5}}{(c_1\cdot\lambda_{10}^{-5}\cdot L_{10}^{-1} + 1)^{(\lambda_{10}\cdot\lambda_{11}^{-1})} - 1} - k\cdot L_{10}$ $\tau_{10/11} = a_0\cdot\omega + a_1$ $\varphi_{10/11} = a_2\cdot\ln(\omega) + a_3$	(Wang et al., 2023)

4. Results

4.1. L9 TIRS-2 calibration results

Statistics of the differences between L9 TIRS-2 L_i and T_i values and those determined from ground LSTs were assessed to obtain systematic (mean error or bias) and random (SD) uncertainties, and root-mean-square differences (RMSD). We also computed three different statistical tests to compare them in order to reinforce the statistical results. First, we used a paired T-test, in which the null hypothesis was that the means of satellite and ground data were equal. The function "ttest2" in Matlab was used, with a 5 % significance level for rejecting the null hypothesis. The second test was the analysis of variance (ANOVA), used to determine whether temperatures and radiances from both groups have a common mean ("anova1" function in Matlab). Finally, we assessed the Levene's test in which the null hypothesis was that both groups had equal variances ("vartestn" function in Matlab using "LeveneAbsolute" option). Table 2 shows bias, SD and RMSD values, and also p-values for the abovementioned tests, for: a) the single value of the pixel covering the site, and b) 3x3 100-m pixel array and c) 5x5 100-m pixel array over the site in L9 TIRS-2 images. Table 2 shows close results for all of them (with negligible differences, < 0.005 W/m^2·sr·μm and < 0.03 K), which proved again the site homogeneity. Results for TIRS-2 OC and RP data are shown in Table 2. Since OC data were no longer provided after the reprocessing in March 2023, they were evaluated from December 2021 to February 2023, with a total of 17 cloud-free matchups of the 23 measurement days in the study period. Results for RP data are shown both for the same period (17 matchups), to permit a proper comparison of calibration results, and from December 2021 to August 2023, with the total of 23 cloud-free matchups in this case.

Fig. 5a shows T_i retrieved from ground data against satellite T_i (average values for the 5x5 100-m pixel array) with OC data for b10 and b11. Fig. 5b shows the same comparison but for RP data. In these figures, errors bars of TIRS-2 TOA T_i show the uncertainties calculated as the square sum of the SD of the pixels in the array and NEΔT/n.pixels$^{1/2}$. These uncertainties ranged from 0.06 to 0.5 K for both bands. Error bars of T_i retrieved from ground data were determined as detailed in Section 3.1.

In Table 2, the p-values of the three statistical hypothesis tests show that the null hypothesis cannot be rejected in any case, giving the high p-values shown. This fact reinforces the hypothesis that the L9 L_i and T_i data and those obtained from ground measurements are equivalent in terms of means and variances. Moreover, the ANOVA F statistics (i.e., between-group variance divided by within-group variance) were close to zero ($\leq$0.01 for all of them).

As shown in Table 2 and in Fig. 5a, TIRS-2 OC L_i and T_i data agreed with those estimated from ground data, with slight negative biases (−0.3 K for b10 and −0.1 K for b11) and RMSDs of 0.5 K for b10 and 0.6 K for b11. The TIRS-2 underestimation was also observed by Barsi et al. (2022) using data measured in inland waters as reference for the L9 data evaluation up to April 2022. In that case, bias was of −0.7 K for b11 and −0.2 K for b10.

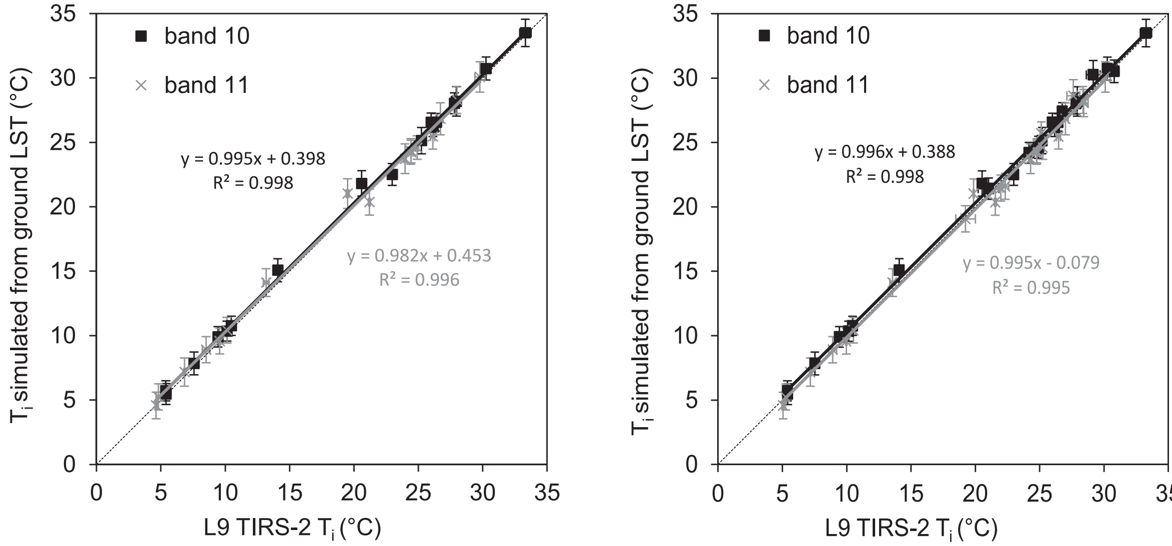
No appreciable T_{10} bias difference (of around −0.01 K) between RP and OC data was shown in Table 2. However, a significant bias difference of 0.37 K in T_{11} was found. Therefore, the 2023 reprocessing introduced a larger calibration change for b11, increasing the difference between biases in b10 and b11 (from −0.17 K for OC data to −0.55 K for RP data), with potential consequences when applying SW algorithms. The last three columns of Table 2 shows the results for RP data with the total number of matchups (23), yielding close results to those for 17 matchups. Like Table 2, Fig. 5b reveals the band difference in terms of biases for RP data, not observed in Fig. 5a for OC data.

Similarly, the evaluation was done in terms of LST. Table 3 shows the validation results for the TIRS-2 retrieved LSTs using the RTE on b10 and b11 data, as explained in Section 3.1, for the 5x5 100-m pixel array. Close results were obtained in terms of LSTs for the 3x3 100-m pixel array and the single pixel covering the site, with negligible differences among them (<0.04 K), similarly to those shown in Table 2. Therefore, the analysis of results is shown only for the 5x5 100-m pixel array hereafter. Fig. 6 shows in situ LSTs against LSTs retrieved from TIRS-2 data.

Table 3 and Fig. 6 show again the good agreement in terms of LSTs estimated using OC data, with slight underestimates of −0.4 K and −0.1 K and RMSDs of 0.6 K and 0.8 K for b10 and b11. A notable disparity in biases is shown between both bands when using RP data (-0.4 K for b10 and 0.4 K for b11).

Table 2
Statistics of the L_i and T_i differences (satellite minus reference data) for the two TIRS-2 bands, both with OC and RP data, for a single pixel (1p) and 3x3 100-m array (3x3) and 5x5 100-m array (5x5).

		OC data (N = 17)			RP data (N = 17)			RP data (N = 23)		
		1p	3x3	5x5	1p	3x3	5x5	1p	3x3	5x5
L_{10} (10^{-2} W/m^2 sr μm)	Bias	−4	−4	−4	−4	−4	−4	−4	−4	−4
	SD	5	5	5	5	5	5	6	6	6
	RMSD	6	7	7	7	7	7	7	7	7
	ANOVA p	0.93	0.93	0.93	0.93	0.92	0.92	0.92	0.92	0.91
	T-test p	0.93	0.93	0.93	0.92	0.92	0.91	0.92	0.92	0.91
	Levene p	0.97	0.98	0.98	0.98	0.98	0.98	0.91	0.92	0.93
L_{11} (10^{-2} W/m^2 sr μm)	Bias	−1	−1	−1	3	2	2	2	2	2
	SD	6	6	6	6	6	7	7	7	7
	RMSD	7	7	7	7	7	7	7	7	7
	ANOVA p	0.97	0.97	0.97	0.93	0.94	0.94	0.93	0.93	0.94
	T-test p	0.97	0.96	0.96	0.93	0.93	0.94	0.93	0.93	0.94
	Levene p	0.89	0.90	0.89	0.89	0.90	0.90	0.91	0.92	0.93
T_{10} (K)	Bias	−0.3	−0.3	−0.3	−0.3	−0.3	−0.3	−0.3	−0.3	−0.3
	SD	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
	RMSD	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	ANOVA p	0.93	0.93	0.93	0.92	0.92	0.92	0.92	0.92	0.91
	T-test p	0.93	0.93	0.93	0.92	0.92	0.91	0.92	0.92	0.91
	Levene p	0.96	0.97	0.96	0.97	0.97	0.97	0.91	0.93	0.94
T_{11} (K)	Bias	−0.1	−0.1	−0.1	0.2	0.2	0.2	0.2	0.2	0.2
	SD	0.6	0.6	0.6	0.5	0.5	0.5	0.6	0.6	0.6
	RMSD	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	ANOVA p	0.97	0.96	0.96	0.94	0.94	0.94	0.93	0.93	0.94
	T-test p	0.97	0.96	0.96	0.93	0.93	0.94	0.93	0.93	0.93
	Levene p	0.88	0.89	0.89	0.90	0.91	0.90	0.93	0.94	0.94

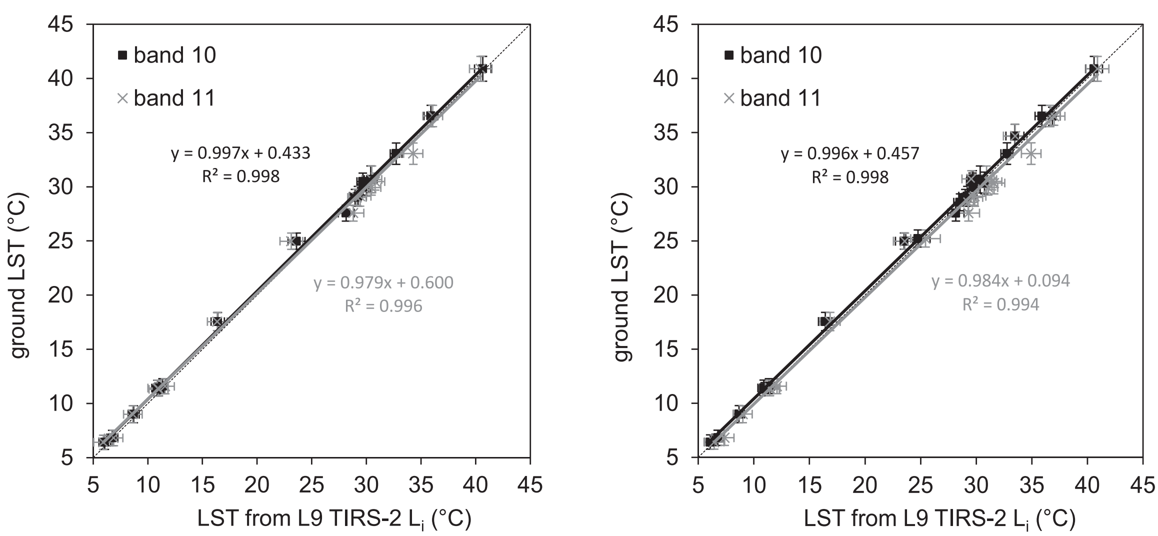


(a) (b)

Fig. 5. T_i retrieved from ground data against L9 TIRS-2 T_i values when using TIRS-2 (a) OC data (17 matchups), and (b) RP data (23 matchups). Linear regressions are shown as solid lines. Regression equations and coefficients of determination are included. The dotted line is the 1:1 identity line.

Table 3
Statistics of the LST differences (satellite minus reference data) when using the RTE both with OC and RP data.

	OC data (N = 17)			RP data (N = 17)			RP data (N = 23)		
Bias	**SD**	**RMSD**	**Bias**	**SD**	**RMSD**	**Bias**	**SD**	**RMSD**	
LST_{10} (K)	−0.4	0.5	0.6	−0.4	0.5	0.6	−0.4	0.5	0.6
LST_{11} (K)	−0.1	0.8	0.8	0.4	0.8	0.9	0.3	0.9	0.9



(a) (b)

Fig. 6. Reference LST versus LSTs retrieved from TIRS-2 L_i with (a) OC data, and (b) RP data.

Table 4
Statistics of the differences between the L2 LST product and ground LST for OC and RP data. The same is shown for the alternative SC_JM_10 and SC_JM_11 algorithms.

	OC data (N = 17)			RP data (N = 17)			RP data (N = 23)		
Bias	**SD**	**RMSD**	**Bias**	**SD**	**RMSD**	**Bias**	**SD**	**RMSD**	
LST_{L2} (K)	0.4	1.0	1.1	0.4	1.0	1.1	0.5	1.1	1.2
LST_{SC_JM10} (K)	−0.7	0.9	1.2	−0.7	1.0	1.2	−0.9	1.1	1.4
LST_{SC_JM11} (K)	1.1	1.9	2.2	1.6	2.0	2.5	1.0	2.2	2.5

4.2. Validation results for the L2 LST product and alternative SC algorithms

Table 4 includes validation results for the L2 LST product generated operationally from TIRS-2 b10 data, and the SC algorithm proposed in Jiménez-Muñoz at al. (2014) applied to TIRS-2 data in b10 and b11 separately. Fig. 7 plots the comparison of the ground LSTs with L2 LSTs (from OC and RP data). X-axis error bars shows the square sum of the average uncertainty for the LSTs over the 5 x5 100-m pixel array and the LST spatial variability. L2 product provides the LST uncertainty for each pixel, which includes the different error sources (those from emissivity, atmospheric parameters, etc.) and the nearest cloud distance effect (Malakar et al., 2018). For the latter, Laraby et al. (2016) studied the relationship between cloud proximity and LST errors reported using buoy data as reference. We observed that the cloud distance effect increases the LST uncertainty from around 2 K (for a cloud distance > 10 km) to more than 5 K (for a cloud distance < 1 km). Some cases in our data set were affected by the latter condition (see Fig. 7), while the ground measurements were carried out only for cloud-free conditions, at least 10 km away from the site. Thus, LST product uncertainties were largely overestimated for some cases in our site because of wrong cloud detection.

Table 4 shows a slight and similar LST overestimation of 0.4–0.5 K for the L2 product both with OC and RP data, with the overestimation increasing with temperature (Fig. 7). However, the results shown in Table 3 for the RTE LST for b10 reveal an underestimation (−0.4 K). The validation results for the SC_JM_10 algorithm applied to b10, with ε_{10} values measured at the site and W estimated for each L9 scene, show also an underestimation (of around −0.7 K).

The emissivities included in the L2 product were close to those measured at the site for flooded soil and bare soil conditions (with differences up to 0.003). However, this was not the case for the full vegetation conditions, the emissivities in the product being lower than those measured at the site by 0.010–0.013. The emissivity underestimation explains in part the systematic positive bias (even larger than 1 K) observed for full vegetation cover, which skew the overall L2 results to a positive bias. Additionally, we observed unexpected too low atmospheric downwelling radiance values in the L2 product, which are of around 50 % of the upwelling radiances included. This fact could also explain in part the positive bias.

In any case, RMSDs of 1.1–1.2 K were shown for the L2 product, which are much lower than those obtained in Meng et al. (2022) and Ye et al. (2022) with pyrgeometer data. On the other hand, the alternative SC_JM_10 and SC_JM_11 algorithms provided larger uncertainties than the RTE (Table 3). This fact was expected, as mentioned in Section 3.2, according to the findings of Jiménez-Muñoz et al. (2014) and Yu et al. (2014). The alternative algorithms also provided larger uncertainties than the L2 LST product (Table 4). Furthermore, results for the SC_JM_11, with coefficients given in Guo et al. (2020), showed RMSD 1 K larger than those for the SC_JM_10, and thus the use of SC_JM_11 is not recommended.

4.3. Validation results for the SW algorithms

Table 5 summarizes the validations corresponding to the 6 types of SW algorithms, and their variants (Section 3.3), when they were applied to OC and RP T_{10} and T_{11} data. The same ground LST data were used as reference for the analysis of results. Results for RP data are given for the total of 23 matchups, with close results for 17 matchups. Fig. 8 shows a box plot comparing the results for the different algorithms.

Table 5
Statistics of the LST differences (satellite minus ground data) obtained for the SW techniques when using TIRS-2 T_{10} and T_{11} as input variables, both from OC and RP data. See Section 3.3 for the description of the SW versions.

SW equation	SW LST – ground LST (K)					
	OC data (N = 17)			RP data (N = 23)		
	Bias	SD	RMSD	Bias	SD	RMSD
SW1	−0.7	0.7	1.0	−1.3	0.7	1.4
SW1_M_W	−0.4	0.7	0.8	−1.0	0.8	1.2
SW1_JM_M	−0.4	0.9	1.0	−1.0	0.9	1.3
SW2_M_W	−0.3	0.7	0.7	−0.8	0.7	1.1
SW2_M	−0.1	0.7	0.7	−0.6	0.6	0.9
SW2_D_W	0.8	0.8	1.1	0.2	0.8	0.8
SW2_D	0.8	0.9	1.2	0.3	0.8	0.9
SW2_D_W_T	0.0	0.8	0.8	−0.6	0.8	1.0
SW2_G	−0.3	0.7	0.8	−0.9	0.7	1.1
SW3_M_W	−0.3	0.7	0.7	−0.9	0.7	1.1
SW3_M	−0.2	0.8	0.8	−0.8	0.9	1.2
SW4_SIM	0.1	0.7	0.7	−0.5	0.7	0.9
SW4_US	0.0	0.6	0.6	−0.6	0.7	0.9
SW4_MLS	0.9	1.3	1.6	0.2	1.3	1.3
SW5_SIM	0.1	0.7	0.7	−0.5	0.7	0.9
SW5_US	−0.2	0.7	0.7	−0.7	0.7	1.0
SW5_MLS	−0.3	0.7	0.8	−0.9	0.7	1.1
SW6_SIM	−0.3	0.6	0.6	−0.8	0.7	1.0
SW6_REG	−1.3	0.7	1.5	−1.8	0.7	2.0

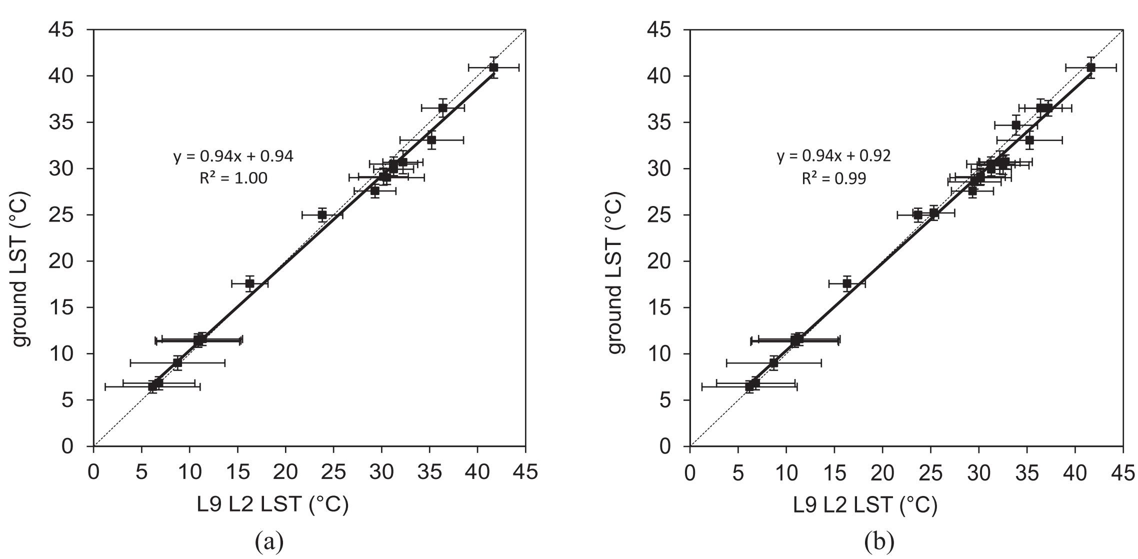
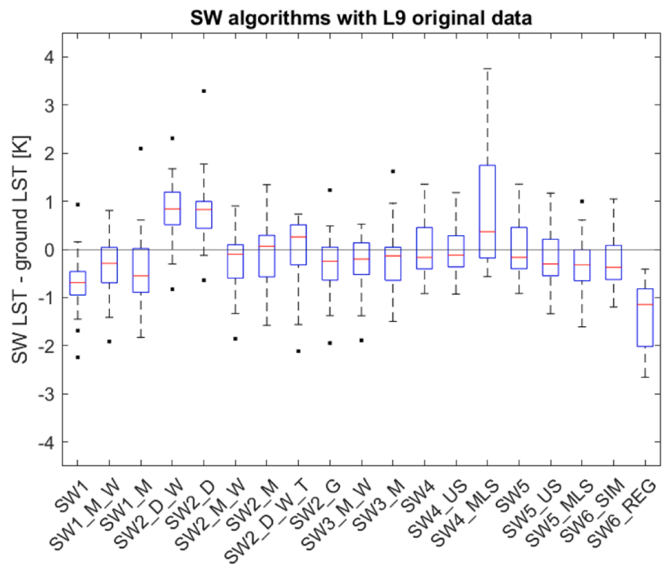
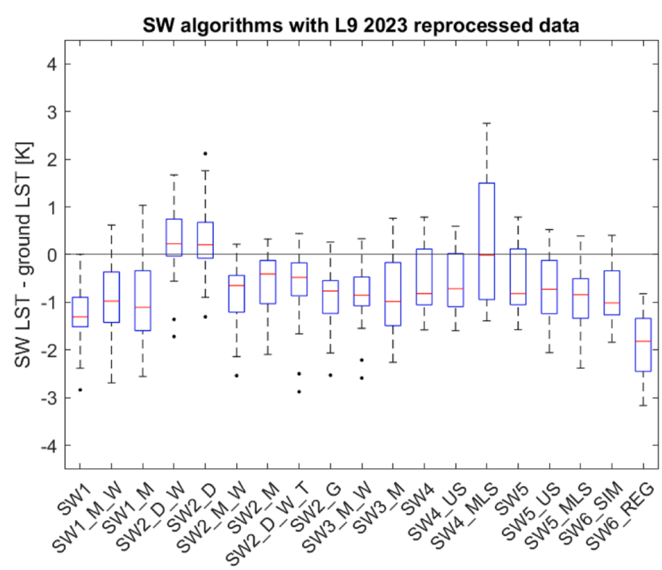


Fig. 7. Ground LSTs against L2 LSTs when using (a) OC data, and (b) RP data.



(a)



(b)

Fig. 8. Comparison of SW algorithm validation for the TIRS-2 (a) OC data and (b) RP data. Box plots show median values, and first and third quartiles. Outliers are shown as dots. Whiskers corresponds to approximately 99.3% coverage.

Table 5 and Fig. 8 show an overall difference of around −0.6 K in the biases obtained for the SW algorithms between RP and OC data, which leads to larger negative biases for most SW algorithms with RP data. When using OC data, biases varied from −1.3 K to 0.9 K depending on the algorithm, with an average of the biases of −0.08 K. Biases from −1.8 K to 0.3 K were shown with RP data, and the average was −0.7 K. However, SDs are quite similar for OC and RP data. The reprocessing increases the difference between the systematic uncertainties in T_{10} and T_{11} (from −0.17 K in OC data to −0.55 K in RP data, Table 2), and leads to a significant change in the biases obtained when using the SW algorithms.

According to the results when using OC data, several SW algorithms provided RMSDs ≤ 1 K, such as different versions of the algorithm based on that proposed by Wan and Dozier (1996) and Wan (2014) (i.e., SW2_M_W, SW2_M, SW2_D_W_T, and SW2_G). Additionally, RMSDs ≤ 1 K were also found for the rest of algorithm types, except in the case of using τ_i regressions on W for Midlatitude atmospheres in SW4_MLS and in the case of using τ_i and φ_i regressions on W in SW6_REG.

Due to the change in T_{11} bias introduced by the 2023 reprocessing, the RMSDs of most of them exceed 1 K with RP data. Bias increased from −0.3 K to −0.9 K for the SW2_G proposed for the future operational SW LST product, and thus RMSD increased from 0.8 K to 1.1 K. After the 2023 reprocessing, the SW algorithm with the lowest RMSD is the SW2_D_W version.

In any case, biases and SDs for most of the SW algorithms studied, both applied to OC and RP data, are within the recommended threshold of 1 K established by the GCOS for LST satellite-retrieved products (GCOS, 2016).

5. Summary and conclusions

This paper provides, first, vicarious calibrations of the L9 TIRS-2 radiances and brightness temperatures in b10 and b11 against ground TIR radiance data measured in a thermally homogeneous site using multi-instrument, spatial and temporal sampling along transects. The vicarious calibrations were performed using L1 data from the original calibration and after the March 2023 reprocessing (OC and RP data, respectively), in which absolute radiances were updated (with the largest adjustment in b11) and some detectors were replaced in both bands to reduce striping.

According to our evaluations at the Valencia test site, close results were obtained for OC and RP data for b10, with radiance biases (RMSDs) of −0.04 (0.07) W/m^2·sr·μm, and brightness temperature biases (RMSDs) of −0.3 (0.5) K. However, results for b11 were different for OC and RP data. While a negligible underestimation was shown for OC data (with radiance bias (RMSD) of −0.01 (0.07) W/m^2·sr·μm, and −0.1 (0.6) K in brightness temperatures), a slight overestimation was observed with RP data (of 0.02 (0.07) W/m^2·sr·μm and 0.2 (0.6) K, respectively). Therefore, the 2023 reprocessing led to an increase in the difference between the systematic uncertainties in T_{10} and T_{11}, from −0.17 K in OC data to −0.55 K in RP data, which was reflected in the results of SW algorithms. Thus, based on our results, the change introduced in the calibration of b11 by the reprocessing seems to be overestimated, accordingly to the low T_{11} bias obtained for OC data at the site.

Secondly, we evaluated the results in LSTs when using the RTE on the L9 TIRS-2 radiances in b10 and b11, i.e., a SC method per band, with emissivity values measured at the site and simulated atmospheric parameters. In that case, LST biases (RMSDs) of −0.4 (0.6) K were obtained when using b10 data, both for OC and RP data. However, for b11, a bias (RMSD) of −0.1 (0.8) K was obtained with OC data, but they were 0.3 K (0.9 K) with RP data. In contrast, positive biases of 0.4–0.5 K, and RMSDs of 1.1–1.2 K, were obtained when validating the L2 product generated operationally from b10 data in C2. An emissivity underestimation detected in the L2 product at the site in full vegetation cover conditions could explain, in part, the systematic positive bias. Additionally, unexpected too low atmospheric downwelling radiances were observed in the L2 product at the site, and this fact could also contribute to the positive bias. However, the results of the SC algorithm proposed in Jiménez-Muñoz et al. (2014) applied to b10 data gave a negative bias from −0.7 K to −0.9 K, with RMSD from 1.2 K to 1.4 K.

Finally, the validations for most of the SW equations tested for L9 data showed biases (accuracy) and SDs (precision) within the recommended GCOS threshold of 1 K, for both OC and RP data, being one of them the SW algorithm provided in Gerace et al. (2020) (SW2_G) for a new operational LST product in the next C3. However, a significant systematic difference (about −0.6 K) was observed between biases when using RP data and biases for OC data, due to the change in brightness temperature systematic uncertainties between both bands with the reprocessing. Thus, systematic uncertainties for most of the SW algorithms, like that proposed in Gerace et al. (2020), were increased with the reprocessing, and RMSDs also increased accordingly.

In any case, the CAL/VAL results reported here for L9 TIRS-2 OC data and the current RP data show good accuracies and precisions to monitor accurately the LST at the high spatial resolution provided by the sensor and to use it for the applications mentioned in Section 1. No significant calibration errors were observed, unlike in the first calibrations of L8 TIRS due to the stray light effect, and both SC and SW methods can provide LSTs from L9 TIRS-2 data with uncertainties around 1 K.

The methodology of using multi-instrument spatial and temporal TIR sampling along transects provided highly accurate and representative ground data. These data yielded high quality CAL/VAL results for L9 TIRS-2 at the Valencia test site. However, more ground data will be acquired following the methodology to allow evaluating possible further reprocessings of TIRS-2 data. Next high-resolution thermal missions will also require high-quality data such as these, and thus the methodology will be applicable at the site for them and it could be extended to other sites for CAL/VAL purposes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was carried out in the framework of the project PID2020-118797RB-I00 (Tool4Extreme) funded by the Spanish Ministry of Science and Innovation and the Spanish State Research Agency (MCIN/AEI/10.13039/501100011033). We also acknowledge the project PROMETEO/2021/016 funded by Generalitat Valenciana. Authors would like to thank the useful information about the L9 status provided by Dr. J. Barsi, from NPL-NASA.

References

Anderson, M.C., Allen, R.G., Morse, A., Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. Remote Sens. Environ. 122, 50–65. https://doi.org/10.1016/j.rse.2011.08.025.

Barsi, J.A., Montanaro, M., Thome, K., Raqueno, N.G., Hook, S.J., Anderson, C., Micijevic, E., 2022. Early radiometric performance of Landsat-9 Thermal Infrared Sensor, in: Butler, J.J., Xiong, X. (Jack), Gu, X. (Eds.), Earth Observing Systems XXVII. SPIE, p. 31. 10.1117/12.2634058.

Berk, A., Anderson, G.P., Acharya, P.K., Shettle, E.P., 2008. MODTRAN5. 2.0. 0 user's manual. Spectr. Sci. Inc., Burlingt. MA, Air Force Res. Lab. Hanscom MA.

Brogniez, G., Pietras, C., Legrand, M., Dubuisson, P., Haeffelin, M., 2003. A High-Accuracy Multiwavelength Radiometer for In Situ Measurements in the Thermal Infrared. Part II: Behavior in Field Experiments. J. Atmos. Ocean. Technol. 20, 1023–1033. https://doi.org/10.1175/1520-0426(2003)20<1023:AHMRFI>2.0.CO;2.

Chatterjee, R.S., Singh, N., Thapa, S., Sharma, D., Kumar, D., 2017. Retrieval of land surface temperature (LST) from landsat TM6 and TIRS data by single channel radiative transfer algorithm using satellite and ground-based inputs. Int. J. Appl. Earth Obs. Geoinf. 58, 264–277. https://doi.org/10.1016/j.jag.2017.02.017.

Coll, C., Caselles, V., Galve, J.M., Valor, E., Niclòs, R., Sánchez, J.M., Rivas, R., 2005. Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data. Remote Sens. Environ. 97, 288–300. https://doi.org/10.1016/j.rse.2005.05.007.

Coll, C., Caselles, V., Valor, E., Niclòs, R., 2012. Comparison between different sources of atmospheric profiles for land surface temperature retrieval from single channel thermal infrared data. Remote Sens. Environ. 117, 199–210. https://doi.org/10.1016/j.rse.2011.09.018.

Coll, C., Niclòs, R., Puchades, J., García-Santos, V., Galve, J.M., Pérez-Planells, L., Valor, E., Theocharous, E., 2019. Laboratory calibration and field measurement of land surface temperature and emissivity using thermal infrared multiband radiometers. Int. J. Appl. Earth Obs. Geoinf. 78, 227–239. https://doi.org/10.1016/j.jag.2019.02.002.

Darge, Y.M., Hailu, B.T., Muluneh, A.A., Kidane, T., 2019. Detection of geothermal anomalies using Landsat 8 TIRS data in Tulu Moye geothermal prospect, Main Ethiopian Rift. Int. J. Appl. Earth Obs. Geoinf. 74, 16–26. https://doi.org/10.1016/j.jag.2018.08.027.

Du, C., Ren, H., Qin, Q., Meng, J., Zhao, S., 2015. A Practical Split-Window Algorithm for Estimating Land Surface Temperature from Landsat 8 Data. Remote Sens. (Basel) 7, 647–665. https://doi.org/10.3390/rs70100647.

García-Llamas, P., Suárez-Seoane, S., Fernández-Guisuraga, J.M., Fernández-García, V., Fernández-Manso, A., Quintano, C., Taboada, A., Marcos, E., Calvo, L., 2019. Evaluation and comparison of Landsat 8, Sentinel-2 and Deimos-1 remote sensing indices for assessing burn severity in Mediterranean fire-prone ecosystems. Int. J. Appl. Earth Obs. Geoinf. 80, 137–144. https://doi.org/10.1016/j.jag.2019.04.006.

GCOS, 2016. The Global Observing System for Climate: Implementation Needs. World Meteorol. Organ. 200, 341.

Gerace, A., Kleynhans, T., Eon, R., Montanaro, M., 2020. Towards an operational, split window-derived surface temperature product for the thermal infrared sensors onboard Landsat 8 and 9. Remote Sens. 12 https://doi.org/10.3390/rs12020224.

Ghent, D., Veal, K., Trent, T., Dodd, E., Sembhi, H., Remedios, J., 2019. A new approach to defining uncertainties for MODIS land surface temperature. Remote Sens. 11 https://doi.org/10.3390/rs11091021.

Good, E.J., Aldred, F.M., Ghent, D.J., Veal, K.L., Jimenez, C., 2022. An Analysis of the Stability and Trends in the LST_cci Land Surface Temperature Datasets Over Europe. Earth Sp. Sci. 9 https://doi.org/10.1029/2022EA002317.

Guillevic, P.C., Privette, J.L., Coudert, B., Palecki, M.A., Demarty, J., Ottlé, C., Augustine, J.A., 2012. Land surface temperature product validation using NOAA's surface climate observation networks-scaling methodology for the visible infrared imager radiometer suite (VIIRS). Remote Sens. Environ. 124, 282–298. https://doi.org/10.1016/j.rse.2012.05.004.

Guo, J., Ren, H., Zheng, Y., Lu, S., Dong, J., 2020. Evaluation of land surface temperature retrieval from landsat 8/TIRS images before and after stray light correction using the SURFRAD dataset. Remote Sens. 12 https://doi.org/10.3390/rs12061023.

Hulley, G.C., Hook, S.J., 2009. The North American ASTER Land Surface Emissivity Database (NAALSED) Version 2.0. Remote Sens. Environ. 113, 1967–1975. https://doi.org/10.1016/j.rse.2009.05.005.

Hulley, G.C., Gottsche, F.M., Rivera, G., Hook, S.J., Freepartner, R.J., Martin, M.A., Cawse-Nicholson, K., Johnson, W.R., 2021. Validation and Quality Assessment of the ECOSTRESS Level-2 Land Surface Temperature and Emissivity Product. IEEE Trans. Geosci. Remote Sens. 1–23 https://doi.org/10.1109/TGRS.2021.3079879.

Jimenez-Munoz, J.C., Sobrino, J.A., Skokovic, D., Mattar, C., Cristobal, J., 2014. Land surface temperature retrieval methods from landsat-8 thermal infrared sensor data. IEEE Geosci. Remote Sens. Lett. 11, 1840–1843. https://doi.org/10.1109/LGRS.2014.2312032.

Krishnan, P., Meyers, T.P., Hook, S.J., Heuer, M., Senn, D., Dumas, E.J., 2020. Intercomparison of in situ sensors for ground-based land surface temperature measurements. Sensors (switzerland) 20, 1–26. https://doi.org/10.3390/s20185268.

Lagouarde, J.-P., Irvine, M., Guillevic, P., 2019. New Insights of Ground-Based Land Surface Temperature Measurements Protocols for Improving Validation of Thermal Infrared Satellite Data, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 9184–9187. 10.1109/IGARSS.2019.8898312.

Laraby, K.G., Schott, J.R., Raqueno, N., 2016. Developing a confidence metric for the Landsat land surface temperature product, in: Velez-Reyes, M., Messinger, D.W. (Eds.). p. 98400C. 10.1117/12.2222582.

Li, Z., Wu, H., Duan, S., Zhao, W., Ren, H., Liu, X., Leng, P., Tang, R., Ye, X., Zhu, J., Sun, Y., Si, M., Liu, M., Li, J., Zhang, X., Shang, G., Tang, B., Yan, G., Zhou, C., 2023. Satellite Remote Sensing of Global Land Surface Temperature: Definition, Methods, Products, and Applications. Rev. Geophys. 61 https://doi.org/10.1029/2022RG000777.

Malakar, N.K., Hulley, G.C., Hook, S.J., Laraby, K., Cook, M., Schott, J.R., 2018. An Operational Land Surface Temperature Product for Landsat Thermal Data: Methodology and Validation. IEEE Trans. Geosci. Remote Sens. 56, 5717–5735. https://doi.org/10.1109/TGRS.2018.2824828.

Meng, X., Cheng, J., Zhao, S., Liu, S., Yao, Y., 2019. Estimating land surface temperature from Landsat-8 data using the NOAA JPSS enterprise algorithm. Remote Sens. 11 https://doi.org/10.3390/rs11020155.

Meng, X., Cheng, J., Guo, H., Guo, Y., Yao, B., 2022. Accuracy Evaluation of the Landsat 9 Land Surface Temperature Product. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 8694–8703. https://doi.org/10.1109/JSTARS.2022.3212736.

Mira, M., Valor, E., Caselles, V., Rubio, E., Coll, C., Galve, J.M., Niclòs, R., Sánchez, J.M., Boluda, R., 2010. Soil moisture effect on thermal infrared (813-μm) emissivity. IEEE Trans. Geosci. Remote Sens. 48, 2251–2260. https://doi.org/10.1109/TGRS.2009.2039143.

Montanaro, M., Gerace, A., Rohrbach, S., 2015. Toward an operational stray light correction for the Landsat 8 Thermal Infrared Sensor. Appl. Opt. 54, 3963–3978. https://doi.org/10.1364/AO.54.003963.

Montanaro, M., McCorkel, J., Tveekrem, J., Stauder, J., Mentzell, E., Lunsford, A., Hair, J., Reuter, D., 2022. Landsat 9 Thermal Infrared Sensor 2 (TIRS-2) Stray Light Mitigation and Assessment. IEEE Trans. Geosci. Remote Sens. 60, 1–8. https://doi.org/10.1109/TGRS.2022.3177312.

Niclòs, R., Doña, C., Valor, E., Bisquert, M., 2014. Thermal-infrared spectral and angular characterization of crude oil and seawater emissivities for oil slick identification. IEEE Trans. Geosci. Remote Sens. 52, 5387–5395. https://doi.org/10.1109/TGRS.2013.2288517.

Niclòs, R., Pérez-Planells, L., Coll, C., Valiente, J.A., Valor, E., 2018. 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. 135, 1–12. https://doi.org/10.1016/j.isprsjprs.2017.10.017.

Niclòs, R., Puchades, J., Coll, C., Barberà, M.J., Pérez-Planells, L., Valiente, J.A., Sánchez, J.M., Juan, M.S., 2021. 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. 174, 237–253. https://doi.org/10.1016/j.isprsjprs.2021.02.005.
Pearlman, A., Efremova, B., Montanaro, M., Lunsford, A., Reuter, D., McCorkel, J., 2022. Landsat 9 Thermal Infrared Sensor 2 On-Orbit Calibration and Initial Performance. IEEE Trans. Geosci. Remote Sens. 60, 1–8. https://doi.org/10.1109/TGRS.2022.3183551.
Peres, L. de F., Lucena, A.J. de, Rotunno Filho, O.C., França, J.R. de A., 2018. The urban heat island in Rio de Janeiro, Brazil, in the last 30 years using remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 64, 104–116. 10.1016/j.jag.2017.08.012.
Qin, Z., Karnieli, A., Berliner, P., 2001. A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel-Egypt border region. Int J Remote Sens 22, 3719–3746. https://doi.org/10.1080/01431160010006971.
Roy, P., Guha, A., Kumar, K.V., 2015. An approach of surface coal fire detection from ASTER and Landsat-8 thermal data: Jharia coal field, India. Int. J. Appl. Earth Obs. Geoinf. 39, 120–127. https://doi.org/10.1016/j.jag.2015.03.009.
Rozenstein, O., Qin, Z., Derimian, Y., Karnieli, A., 2014. Derivation of Land Surface Temperature for Landsat-8 TIRS Using a Split Window Algorithm. Sensors 14, 5768–5780. https://doi.org/10.3390/s140405768.
Vidal, T.H.G., Gamet, P., Olioso, A., Jacob, F., 2022. Optimizing TRISHNA TIR channels configuration for improved land surface temperature and emissivity measurements. Remote Sens. Environ. 272, 112939 https://doi.org/10.1016/j.rse.2022.112939.
Wan, Z., 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sens. Environ. 140, 36–45. https://doi.org/10.1016/j.rse.201.
Wan, Z., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE Trans. Geosci. Remote Sens. 34, 892–905. https://doi.org/10.1109/36.5084063.08.027.
Wang, M., He, G., Zhang, Z., Wang, G., Wang, Z., Yin, R., Cui, S., Wu, Z., Cao, X., 2019. A radiance-based split-window algorithm for land surface temperature retrieval: Theory and application to MODIS data. Int. J. Appl. Earth Obs. Geoinf. 76, 204–217. https://doi.org/10.1016/j.jag.2018.11.015.
Wang, M., Li, M., Zhang, Z., Hu, T., He, G., Zhang, Z., Wang, G., Li, H., Tan, J., Liu, X., 2023. Land Surface Temperature Retrieval From Landsat 9 TIRS-2 Data Using Radiance-Based Split-Window Algorithm. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 1100–1112. https://doi.org/10.1109/JSTARS.2022.3232621.
Yamada, Y., Harris, S., Hayes, M., Simpson, R., Wimmer, W., Holmes, R., Nightingale, T., Lee, A., Jepsen, N., Morgan, N., Göttsche, F.-M., Niclòs, R., Perelló, M., Donlon, C., Fox, N., 2023a. Results from CEOS International Thermal Infrared Radiometer Inter-comparison (CRIC) Part 2 of 3: Laboratory Comparison of Radiometers.
Yamada, Y., Harris, S., Hayes, M., Simpson, R., Wimmer, W., Holmes, R., Nightingale, T., Lee, A., Jepsen, N., Morgan, N., Göttsche, F.-M., Niclòs, R., Perelló, M., Donlon, C., Fox, N., 2023. CEOS International Thermal Infrared RadiometerComparison: Part I: Laboratory Comparison of Radiometers and Blackbodies. Submitted J. Atmos. Oceanic Technol.
Ye, X., Ren, H., Zhu, J., Fan, W., Qin, Q., 2022. Split-Window Algorithm for Land Surface Temperature Retrieval From Landsat-9 Remote Sensing Images. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2022.3184980.
Yu, X., Guo, X., Wu, Z., 2014. Land Surface Temperature Retrieval from Landsat 8 TIRS—Comparison between Radiative Transfer Equation-Based Method, Split Window Algorithm and Single Channel Method. Remote Sens. 6, 9829–9852. https://doi.org/10.3390/rs6109829.