Continental-scale evaluation of three ECOSTRESS land surface temperature products over Europe and Africa: Temperature-based validation and cross-satellite comparison

Tian Hu, Kaniska Mallick, Glynn C Hulley, Lluís Perez Planells, Frank M Göttsche, Martin Schlerf, Patrik Hitzelberger, Yoanne Didry, Zoltan Szantoi, Itziar Alonso, et al.

Remote Sensing of Environment

Continental evaluation of three ECOSTRESS land surface temperature products over Europe and Africa: Temperature-based validation and cross-satellite comparison with ASTER and Landsat

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Section/Category:	Land surface reflectance and temperature (including sensor calibration studies)
Keywords:	High spatial resolution LST; ECOSTRESS; ASTER; Landsat
Corresponding Author:	Tian Hu, Ph.D. Luxembourg Institute of Science and Technology Belvaux, Sanem LUXEMBOURG
First Author:	Tian Hu, Ph.D.
Order of Authors:	Tian Hu, Ph.D.
	Kaniska Mallick, Ph.D.
	Glynn C. Hulley, Ph.D.
	Lluís Perez Planells, Ph.D.
	Frank M. Göttsche, Ph.D.
	Martin Schlerf, Ph.D.
	Patrik Hitzelberger
	Yoanne Didry
	Zoltan Szantoi, Ph.D.
	Itziar Alonso
	José A. Sobrino, Ph.D.
	Dražen Skoković, Ph.D.
	Jean-Louis Roujean, Ph.D.
	Gilles Boulet, Ph.D.
	Philippe Gamet, Ph.D.
	Simon Hook, Ph.D.
Abstract:	High spatial resolution land surface temperature (LST, <100 m) is used in a wide range of applications such as agricultural water consumptive use estimation, crop water stress monitoring, fire mapping, urban heat island and volcano eruption detection. LST retrievals from the ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) launched in June 2018, together with the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER, launched in 1999) and the Landsat series, comprise the state-of-the-art high spatial resolution LST datasets publicly accessible. The U.S. Geological Survey (USGS) Landsat LST is generated using the single-channel (SC) algorithm. The ASTER and ECOSTRESS LST are generated using the temperature and emissivity separation (TES) algorithm developed by Jet Propulsion Laboratory (JPL). Recently, we generated the ECOSTRESS LST product over Europe and Africa using both the TES and split-window (SW) algorithms under the European ECOSTRESS Hub (EEH). Although extensive validations have been conducted for the ASTER and Landsat LST products, a further validation of different types of ECOSTRESS LST products is still in need. Especially, a cross-satellite comparison of ECOSTRESS, ASTER and Landsat LST is required for a thorough understanding of the consistency among the different LST products and will furthermore support the adaptation of LST retrieval algorithms for

	future thermal missions. Here, we validated the JPL TES (Collection 1), EEH TES and EEH SW ECOSTRESS LST products over Europe and Africa for the period between August 1, 2018 and December 31, 2021 by comparing against the in-situ measurements at 9 sites over a wide variety of land cover types. Meanwhile, the validation results were compared with those obtained for ASTER and Landsat LST at the same sites. The results reveal that the three ECOSTRESS LST products have consistent performances, with an RMSE around 2 K overall. A cold bias around 1 K exists for all three ECOSTRESS LST, which is related to an issue with the sensor's radiometric calibration in Collection 1 data that is addressed in Collection 2 and to be released in 2022. The Landsat LST shows a similar accuracy, with an RMSE of 2.20 K and bias of 0.54 K. The EEHSW LST show the highest consistency with Landsat LST, probably due to the identical emissivity correction process. The performance of ASTER LST is also similar, with an RMSE of 1.98 K and bias of 0.9 K. σ (precision) of all the LST products are around 1.5 K. Future recalibration of the ECOSTRESS Level 1 radiance data in Collection 2 release is expected to further improve the accuracy of ECOSTRESS LST.
Suggested Reviewers:	Sibo Duan, Ph.D. Professor, Chinese Academy of Agricultural Sciences duansibo@caas.cn Prof. Duan is expertised in thermal remote sensing and validation of land surface temperature and emissivity.
	Michael Abrams, Ph.D. Research Scientist, Jet Propulsion Laboratory mjabrams@jpl.nasa.gov Dr. Abrams has been the US ASTER Science Team Leader since 2003. His research interests include applications of remote sensing data for volcano monitoring and eruption prediction and the development of new instruments for Earth observation.
	Yunyue Bob Yu, Ph.D. Research Scientist, National Oceanic and Atmospheric Administration Boulder yunyue.yu@noaa.gov Dr. Yu is experienced in LST retrieval and validation. He is chairman of land surface application team of the U.S. GOES-R Algorithm Working Group, and the LST and Albedo EDR lead of the U.S. Joint Polar-orbiting Satellite System (JPSS) program.
	Darren Ghent, Ph.D. Research Scientist, University of Leicester djg20@le.ac.uk Dr. Ghent is experienced in LST retrieval and validation. He is the PI of Sentinel-3 LST genetation and ESA CCI LST.
Opposed Reviewers:	Isabel Franco Trigo, Ph.D. Research Scientist, Institudo Dom Luiz isabel.trigo@ipma.pt Conflict of interest
	Sofia Ermida, Ph.D. Research Scientist, Institudo Dom Luiz snermida@fc.ul.pt Conflict of interest

Continental evaluation of three ECOSTRESS land surface temperature products over Europe and Africa: Temperature-based validation and cross-satellite comparison with ASTER and Landsat

Tian Hu[a], Kaniska Mallick[a], Glynn C. Hulley[b], Lluís Perez Planells[c], Frank M. Göttsche[c], Martin Schlerf[a], Patrik Hitzelberger[d], Yoanne Didry[d], Zoltan Szantoi[e,f], Itziar Alonso[e], José A. Sobrino[g], Dražen Skoković[g], Jean-Louis Roujean[h], Gilles Boulet[h], Philippe Gamet[h], Simon Hook[b]

a. Department of Environment Research and Innovation, Luxembourg Institute of Science and Technology, Belvaux 4362, Luxembourg

b. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

c. Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany

d. Department of IT for Innovative Services, Luxembourg Institute of Science and Technology, Belvaux 4362, Luxembourg

e. Science, Applications & Climate Department, European Space Agency, Frascati 00044, Italy

f. Stellenbosch University, Stellenbosch 7602, South Africa

g. Global Change Unit, Imaging Processing Laboratory, University of Valencia, Paterna 46980, Spain

h. Centre d'Etudes Spatiales de la Biosphère, CNES, CNRS, INRA, IRD, UPS, Toulouse 31401, France

Abstract

High spatial resolution land surface temperature (LST, <100 m) is used in a wide range of applications such as agricultural water consumptive use estimation, crop water stress monitoring, fire mapping, urban heat island and volcano eruption detection. LST retrievals from the ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) launched in June 2018, together with the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER, launched in 1999) and the Landsat series, comprise the state-of-the-art high spatial resolution LST datasets publicly accessible. The U.S. Geological Survey (USGS) Landsat LST is generated using the single-channel (SC) algorithm. The ASTER and ECOSTRESS LST are generated using the temperature and emissivity

separation (TES) algorithm developed by Jet Propulsion Laboratory (JPL). Recently, we generated the ECOSTRESS LST product over Europe and Africa using both the TES and split-window (SW) algorithms under the European ECOSTRESS Hub (EEH). Although extensive validations have been conducted for the ASTER and Landsat LST products, a further validation of different types of ECOSTRESS LST products is still in need. Especially, a cross-satellite comparison of ECOSTRESS, ASTER and Landsat LST is required for a thorough understanding of the consistency among the different LST products and will furthermore support the adaptation of LST retrieval algorithms for future thermal missions. Here, we validated the JPL TES (Collection 1), EEH TES and EEH SW ECOSTRESS LST products over Europe and Africa for the period between August 1, 2018 and December 31, 2021 by comparing against the *in-situ* measurements at 9 sites over a wide variety of land cover types. Meanwhile, the validation results were compared with those obtained for ASTER and Landsat LST at the same sites. The results reveal that the three ECOSTRESS LST products have consistent performances, with an RMSE around 2 K overall. A cold bias around 1 K exists for all three ECOSTRESS LST, which is related to an issue with the sensor's radiometric calibration in Collection 1 data that is addressed in Collection 2 and to be released in 2022. The Landsat LST shows a similar accuracy, with an RMSE of 2.20 K and bias of 0.54 K. The EEHSW LST show the highest consistency with Landsat LST, probably due to the identical emissivity correction process. The performance of ASTER LST is also similar, with an RMSE of 1.98 K and bias of 0.9 K. σ (precision) of all the LST products are around 1.5 K. Future recalibration of the ECOSTRESS Level 1 radiance data in Collection 2 release is expected to further improve the accuracy of ECOSTRESS LST.

Keywords: High spatial resolution LST, ECOSTRESS, ASTER, Landsat

1. Introduction

Land surface temperature (LST) is an important biophysical variable that controls the magnitude and variability of terrestrial ecosystem processes (Bai et al. 2022; Li et al. 2013; Mallick et al. 2015; Mallick et al. 2014). It has been identified as a critical Earth System Data Record (ESDR) and Essential Climate Variable (ECV) (NASA 2005, 2011), as well as one of the high-priority parameters of the International Geosphere and Biosphere Program (IGBP) (Townshend et al. 1994). LST carries the imprints of surface water stress and is immensely sensitive to evaporative cooling. It is a preeminent variable for studying evaporation and surface-atmosphere exchange (Mallick et al. 2018), climatic and environmental studies (Hulley et al. 2021), surface radiation budget (Hu et al. 2016; Liang et al. 2010), drought monitoring (Anderson et al. 2016; Hu et al. 2020) and urban heat island (Chakraborty et al. 2021).

High spatial resolution LST (<100 m) is especially important for agricultural applications including agricultural consumptive water estimation, irrigation management, crop water stress monitoring since the pixel size is expected to match the individual field size (Anderson et al. 2012; Anderson et al. 2021; Ekinzog et al. 2022). Moreover, for urban environmental studies, high spatial resolution LST is also needed due to the marked spatial heterogeneity of urban canopies (Liu and Zhang 2011; Yuan and Bauer 2007). Also, natural hazard early warning and mitigation such as fire detection and volcanic process monitoring (Guangmeng and Mei 2004; Ramsey and Flynn 2020; Silvestri et al. 2020) require high spatial resolution LST for risk management.

Thermal infrared (TIR) remote sensing is the most straightforward way to retrieve LST at global scales due to the close relation of longwave TIR radiation from the Earth's surface to surface temperature and the extensive spatial coverage of satellite observations (Becker and Li 1995; Li et al. 2013; McMillin 1975). Several key difficulties exist in LST retrieval from TIR radiance measured at the top of atmosphere (TOA): 1) estimation of LST is an underdetermined problem considering there is always one more unknown than the thermal channels (N

77 emissivities + 1 temperature > N TIR channels), 2) accurate atmospheric correction is difficult
78 to implement due to the uncertainties in the atmospheric water content and vertical profile data
79 and the error propagation in the atmospheric radiative transfer models, 3) decoupling LST and
80 emissivity is challenging considering the pronounced variation of emissivity among different
81 land surface types and soil compositions. Based on different assumptions on how to deal with
82 these challenges, multiple algorithms have been developed to retrieve LST depending on the
83 thermal bands available, including single-channel (SC) (Qin et al. 2001), split-window (SW)
84 (McMillin 1975; Wan and Dozier 1996) and temperature and emissivity separation (TES)
85 (Gillespie et al. 1998) algorithms. These algorithms have been successfully applied to different
86 satellite sensors to generate operational LST products, including Landsat (Malakar et al. 2018),
87 Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Abrams 2000),
88 Moderate Resolution Imaging Spectroradiometer (MODIS) (Wan 2014; Wan and Li 1997),
89 Advanced Baseline Imager (ABI) series (Yu et al. 2008) and Visible Infrared Imaging
90 Radiometer Suite (VIIRS) (Islam et al. 2016; Li et al. 2014), among others.

91 Since the launch in June 2018, the ECOsystem Spaceborne Thermal Radiometer
92 Experiment on Space Station (ECOSTRESS), has acquired the most detailed LST images (38
93 m × 69 m) with the most frequent revisits (3-5 days) ever from space (Fisher et al. 2020; Hook
94 et al. 2019; Hulley et al. 2021), compared with the LST retrievals from the comparable satellites
95 in orbit, i.e., ASTER (90 m, 16-day revisit) (Abrams 2000; Abrams et al. 2015) and Landsat
96 series (60 m for Landsat 7 ETM+, 100 m for Landsat 8/9 TIRS/TIRS2, 16-day revisit) (Wulder
97 et al. 2019). Although considerable effort has been put into the validation activities for ASTER
98 (Sabol Jr et al. 2009; Sobrino et al. 2007; Wang and Liang 2009) and Landsat LST products
99 (Malakar et al. 2018; Wang et al. 2019; Wang et al. 2020), a further evaluation of different
100 types of LST products from the newly launched ECOSTRESS mission is still needed (Hulley
101 et al. 2021), especially inter-comparison with ASTER and Landsat LST products, to better

understand their consistency in different surface and atmospheric scenarios and to facilitate the preparation of future TIR missions such as National Aeronautics and Space Administration's (NASA) future Surface Biology and Geology (SBG) mission in 2026 (Cawse-Nicholson et al. 2021), European Space Agency's (ESA) Land Surface Temperature Monitoring (LSTM) mission in 2028 (Koetz et al. 2019) and the Franco-Indian joint Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) mission in 2025 (Lagouarde et al. 2018).

In this study, we evaluated three ECOSTRESS LST products, which are the NASA official product (Collection 1) generated using the Jet Propulsion Laboratory (JPL) TES algorithm and those generated using the SW and TES algorithms from the ECOSTRESS European Hub (EEH). The evaluation was conducted by comparing each ECOSTRESS LST product with the *in-situ* measurements at 9 sites between August 1, 2018 and December 31, 2021 over Europe and Africa. Additionally, the evaluation results were compared with the ASTER LST produced by the ASTER Science Team and the Landsat LST produced by U.S. Geological Survey (USGS) at the same sites. The evaluation is expected to facilitate the use of ECOSTRESS LST products with a better understanding of their accuracy and support the LST retrieval for the future TIR missions.

2. Data

2.1. Satellite LST products

2.1.1. ECOSTRESS LST

ECOSTRESS was launched to the international space station (ISS) on June 29, 2018. The instrument includes a TIR multispectral whiskbroom scanner with five bands between 8 and 12.5 μm, which scans at ±25° and results in a swath width of ~400 km depending on the ISS height. The ISS orbit allows excellent coverages of the selected targets with multiple revisits in the diurnal cycle. Three different ECOSTRESS LST products were used in this evaluation:

the LST estimated using the TES algorithm of NASA JPL (named JPLTES LST hereafter), that using the TES algorithm by EEH (named EEHTES LST hereafter) and that using the SW algorithm by EEH (named EEHSW LST hereafter).

1) JPLTES LST

Initially, a 5-band TES algorithm was implemented on the ECOSTRESS observations to retrieve LST. After May 15, 2019, a 3-band TES algorithm was used to retrieve LST from the ECOSTRESS thermal bands centered at 8.78, 10.49 and 12.09 μm and the other two bands centered at 8.29 and 9.2 μm were discarded during the data transfer process. This is because the primary and secondary mass storage units (MSU-A and -B) failed in flight during 2019 and TIR bands 1 and 3 were no longer downloaded in the new acquisition mode to maximize the download capacity (Hulley et al. 2021). The atmospheric correction was conducted using the atmospheric Radiative Transfer for TOVS (RTTOV) model (Saunders et al. 2018) with inputs of atmospheric water vapor and temperature profiles from the NASA GMAO GEOS5-FP reanalysis product. The GEOS atmospheric profiles depict the atmospheric state at 42 pressure levels and are available in near real time with a spatial resolution of 0.25° in latitude and ~0.33° in longitude every 3 hours. To match the ECOSTRESS observations, spatial and temporal interpolation were conducted on the GEOS5 atmospheric profiles. The three atmospheric parameters from the RTTOV model outputs (i.e., atmosphere upwelling and downwelling radiances and transmissivity) were used to obtain the surface-leaving radiance, which served as an input into the TES algorithm to estimate LST and emissivity simultaneously in an iterative way. The pixels were aggregated from the native resolution (38 m × 69 m) to square pixels (~70 m × 70 m) to facilitate the subsequent usages. More details of the JPL-TES algorithm can be found in the algorithm theoretical basis document (ATBD, https://ecostress.jpl.nasa.gov/data/atbds-summary-table). The ECOSTRESS JPLTES LST

products (ECO2LSTE) were downloaded from the NASA Earthdata Search platform (https://search.earthdata.nasa.gov/search).

2) EEHTES LST

EEH is a project funded by ESA, aiming at developing LST and evapotranspiration (ET) products for Europe and Africa using the high spatial and temporal resolution ECOSTRESS TIR observations. Two different LST algorithms (i.e., TES and SW) and three ET models were implemented in EEH. The unique feature of EEH is that both the LST algorithms are driven by homogenized radiance and environmental datasets, and all the evaporation models are forced by uniform upper boundary and lower boundary conditions. This characteristic enables appropriate comparisons between different models for a wide range of surface, energy, and water availability scenarios. Overall, the EEH will serve as a support to ESA's next generation Copernicus High Priority Candidate LSTM mission.

The EEHTES LST is retrieved in a similar way to the JPLTES LST except for the atmospheric profiles input into RTTOV for estimating the atmospheric parameters. To generate the EEHTES LST, the European Centre for Medium-range Weather Forecasting (ECMWF) ERA5 hourly products on single level and on 37 pressure levels were used to mimic the near-surface and vertical atmospheric conditions, respectively. Mode details of the EEHTES LST retrieval process can be found in the ATBD (http://isp-projects.private.list.lu/eeh/public/datasets).The EEHTES LST products between August 2018 and December 2021 are available on the Food Security-TEP portal (https://foodsecurity-tep.net/).

3) EEHSW LST

The EEHSW algorithm follows the generalized SW method proposed by Wan and Dozier (1996) as below:

$$T_s = b_0 + \left(b_1 + b_2\frac{1-\varepsilon}{\varepsilon} + b_3\frac{\Delta\varepsilon}{\varepsilon^2}\right)\frac{T_{11}+T_{12}}{2} + \left(b_4 + b_5\frac{1-\varepsilon}{\varepsilon} + b_6\frac{\Delta\varepsilon}{\varepsilon^2}\right)\frac{T_{11}-T_{12}}{2} \quad (1)$$

$$+b_7(T_{11}-T_{12})^2$$

where T_s is the LST (in K), T_{11} and T_{12} are at-sensor brightness temperatures (in K) centred around 11 and 12 μm (bands 4 and 5 in the case of ECOSTRESS), respectively, ε is the mean emissivity of the SW channels (bands 4 and 5), $\Delta\varepsilon$ is the emissivity difference between the SW channels, and b_0 to b_7 are the algorithm coefficients.

To obtain the eight coefficients in Eq. 1, a simulation dataset was compiled using the RTTOV 12 atmospheric radiative transfer model and the SeeBor V5.0 atmospheric profile database. The atmospheric transmittance and the atmospheric upwelling radiance are simulated over land for the SW channels using all the 9136 profiles at viewing zenith angle (VZA) from nadir to 35° with an increment of 5° and the atmospheric downwelling radiance at 53°. The numbers of daytime and night-time profiles are 4948 and 4188, respectively. The volumetric water vapor content (WVC) was divided into six subranges with an overlap of 0.5 g/cm^2: [0, 1.5], [1, 2.5], [2, 3.5], [3, 4.5], [4, 5.5] and [5, 7.8]. To maximize the algorithm performance over a wide range of surface conditions, the LST in the simulation was set as follows. For cold atmospheric profiles (T_0 < 280 K), the LST varies from T_0 - 20 K to T_0 + 4 K in steps of 5 K, and for warm atmospheric profiles (T_0 > 280 K), LST varies from T_0 - 5 K to T_0 + 29 K in steps of 5 K. In addition, a total of 81 emissivity spectra were selected from the ECOSTRESS spectral library, including vegetation, water, ice, snow, rock, sand and soil spectra. Finally, for given LST, LSE, and atmospheric parameters, the TOA radiances in ECOSTRESS bands 4 and 5 were simulated, and the brightness temperatures were calculated. Once the simulation database was established, the eight coefficients were determined via statistical regression.

The enterprise LSE estimation method based on the ASTER GED v3 product was used to provide emissivity (Hulley et al. 2015; Malakar et al. 2018; Yu et al. 2017). The bare soil emissivities were first extracted from the ASTER GED and then converted to the ECOSTRESS SW channels. Adjustments over vegetated surfaces and snow/ice covered surfaces were made

by using the vegetation fractional coverage and fraction snow extent data from the Copernicus Global Land Service (CGLS). Moreover, constant emissivities were assigned to pixels marked as water.

After acquiring the emissivity for each pixel in the image and the regression coefficients stratified by VZA and WVC, LST was calculated directly from the TOA radiances for clear-sky scenes screened using cloud mask. The WVC information was obtained from the ERA5 hourly single level data, which was interpolated temporally and spatially to ECOSTRESS observations. Mode details of the EEHSW LST retrieval process can be found in the ATBD (http://isp-projects.private.list.lu/eeh/public/datasets). The EEHSW LST products between August 2018 and December 2021 are also available on the Food Security-TEP portal (https://foodsecurity-tep.net/).

2.1.2. ASTER LST

ASTER is a high spatial resolution radiometer onboard the NASA Earth Observing System (EOS) Terra satellite, which was launched into a sun-synchronous orbit on December 19, 1999. It collects 14 bands from the visible to the thermal wavelengths, among which five bands are in the TIR spectrum with a spatial resolution of 90 m, centered at 8.30, 8.65, 9.05, 10.60 and 11.30 μm, respectively. The ASTER LST product (AST08) was generated using the TES algorithm from the 5 TIR bands by the U.S./Japan ASTER Science team (Gillespie et al. 1998). Previous studies (Sabol Jr et al. 2009; Sobrino et al. 2007; Wang and Liang 2009) showed that the ASTER LST accuracy is within the 1.5 K and the LST estimation error could increase under unusually humid atmosphere over cold surfaces due to the incomplete atmospheric correction (Sabol Jr et al. 2009). The AST08 products were downloaded from the NASA Earthdata Search platform (https://search.earthdata.nasa.gov/search).

2.1.3. Landsat LST

The Landsat mission is a joint NASA/USGS program. It collects the longest record of high spatial resolution TIR data (60−120 m) since the launch of Landsat 4 (July 1982). The Landsat LST produced by USGS was estimated using the operational single-channel (OSC) algorithm developed by Malakar et al. (2018). The three atmospheric parameters were estimated by inputting the atmospheric profile data into the MODTRAN 5.2 model. The emissivity was estimated in the same way as for the EEHSW LST based on the ASTER GED algorithm. The surface emittance was then calculated as follows:

$$B_i(T_s) = \frac{L_{TOA,i} - L_{up,i} - (1-\varepsilon_i)\tau_i L_{down,i}}{\varepsilon_i \tau_i} \quad (2)$$

where T_s is the surface temperature, B is the Planck function, $L_{TOA,i}$ is the TOA radiance in band i, $L_{up,i}$, $L_{down,i}$ and τ_i are atmosphere upwelling, downwelling radiances and transmissivity, respectively, ε_i is the narrow-band emissivity in band i. Finally, LST was calculated from the surface emittance based on a look up table (Malakar et al. 2018). The RMSE of the USGS Landsat LST was reported to be approximately 2.5 K (Wang et al. 2022). The Landsat-7 and -8 LST data were downloaded from the Google Earth Engine (GEE).

2.2. Ground measurements

Ground measurements from five different networks were used to evaluate the high spatial resolution LST, including the Karlsruhe Institute of Technology (KIT) network, the Copernicus network, the Baseline Surface Radiation Network (BSRN), the Global Change Unit (GCU) network and the Integrated Carbon Observation System (ICOS) network. The detailed information of these sites is listed in Table 1. The spatial distribution of the 9 sites is shown in Fig. 1. The pictures of the landscape around the sites are shown in Fig. 2.

The KIT and Copernicus stations were designed to validate LST over relatively homogeneous surfaces. The surface upwelling and downwelling radiances are collected using narrow-band radiometers measuring TIR radiance between 9.6 and 11.5 μm (Göttsche et al. 2016). The radiometers are mounted at heights between 12 and 28 m and measure the radiances

once per minute, which results in fields of view (FOV) between 3 and 14 m^2. Two KIT sites over desert and water surface and one Copernicus site over mixed forests were selected.

The BSRN was set up to provide observations of the best possible quality for short- and long-wave surface radiation fluxes sampled at high frequency (once per minute). The measurements from well calibrated pyrgeometers are expected to provide reference for the validation of satellite-based estimates of the surface radiative fluxes and for the comparison to climate model calculations (Driemel et al. 2018). Here, only one BSRN site CAB was selected considering both upwelling and downwelling radiation fluxes are collected at this site and the accuracy of the measurements has been demonstrated in the previous studies (Trigo et al. 2021).

The GCU sites were set up in Spain for the calibration of TIR sensors and the validation of satellite LST products (Sobrino and Skoković 2016). Thermal radiance measurements are collected for the spectral range between 8 and 14 μm. The measurements are collected every 10 s and averaged to 5 min. Two permanent sites providing long-term observations were selected from the GCU sites.

The ICOS network was developed to produce standardized, high-precision and long-term observations for understanding the carbon cycle and providing necessary information on greenhouse gases. The ICOS sites measure the fluxes of greenhouse gases, living and non-living components as well as drivers (e.g., radiations) for the exchange of greenhouse gases, water and energy between ecosystems and the atmosphere. We selected three sites over different land surface types. The upwelling and downwelling radiations are measured using pyrgeometers and averaged for each half hour.

Table 1 Information of the selected 9 *in-situ* measurement sites.

Site No.	Site location	Site ID	Network	Latitude	Longitude	Surface type	Emissivity
1	Gobabeb wind tower, Namibia	GBB	KIT	23.551° S	15.051° E	Barren/sparsely vegetated	0.940 Hulley et al. (2021)

2	Lake Constance, Germany	CNS	KIT	47.605° N	9.444° E	Water	0.973 Hulley et al. (2021)
3	KIT Forest, Germany	KIT	Copernicus	49.091° N	8.425° E	Mixed forest	0.988 Freitas et al. (2009)
4	Cabauw, Netherlands	CAB	BSRN	51.971° N	4.927° E	Grassland	From ECOSTRESS
5	Fuente Duque, Donana, Spain	FDU	GCU	36.998° N	6.434° W	Marshland	Measurements Sobrino and Skoković (2016)
6	Balsa Blanca, Cabo de Gata, Spain	CDG	GCU	36.939° N	2.034° W	Woody savannas	Measurements Sobrino and Skoković (2016)
7	Fontainebleau-Barbeau, France	FON	ICOS	48.476° N	2.780° E	Deciduous broadleaf forest	From ECOSTRESS
8	Lison, Italy	LSN	ICOS	45.740° N	12.750° E	Cropland	From ECOSTRESS
9	San Rossore 2, Italy	SR2	ICOS	43.732° N	10.291° E	Evergreen needleleaf forest	From ECOSTRESS

Estimating LST from the radiance measurements of the KIT, Copernicus and GCU sites was achieved by inverting the Planck's law as follows:

$$B_i(T_s) = \frac{L_{up,i} - (1-\varepsilon_i)L_{down,i}}{\varepsilon_i} \quad (3)$$

where $L_{up,i}$ is the upwelling longwave radiance measured by the station radiometers, $L_{down,i}$ is the downwelling thermal radiance, which is measured by an additional radiometer for the KIT and Copernicus sites (Duan et al. 2019) and calculated by inputting MOD07 atmospheric profiles into the MODTRAN model for the GCU sites (Sobrino and Skoković 2016), ε_i is the narrow-band emissivity, T_s is the inverted temperature. The band-effective emissivity was obtained using the methods listed in Table 1.

For the longwave radiation measurements from the BSRN and ICOS sites, the LST is estimated by inverting the Stefan-Boltzmann's law as follows:

$$T_s = \sqrt[4]{\frac{R_{up} - (1-\varepsilon_{BB})R_{down}}{\varepsilon_{BB}\sigma}} \quad (4)$$

where R_{up} and R_{down} are the measured upward and downward longwave radiations, respectively, ε_{BB} is the broadband emissivity and σ is the Stefan-Boltzmann constant. The broadband emissivity was estimated from the ECOSTRESS emissivity retrievals in the three thermal bands as follows:

$$\varepsilon_{BB} = 0.3287\varepsilon_2 + 0.3783\varepsilon_4 + 0.3158\varepsilon_5 - 0.0255 \quad (5)$$

where ε_2, ε_4 and ε_5 are the emissivity retrievals in bands 2, 4 and 5, respectively.

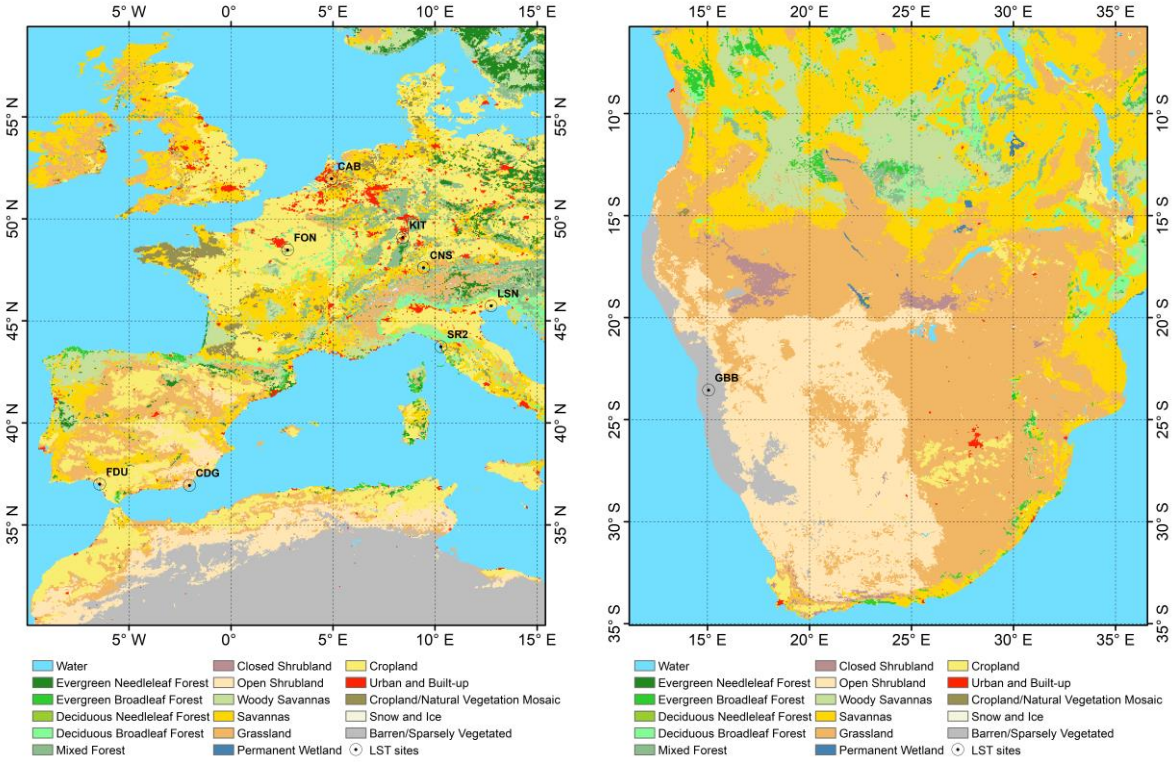


Fig. 1 Spatial distribution of the nine *in-situ* measurement sites, including 8 sites over Europe and 1 site over Africa.

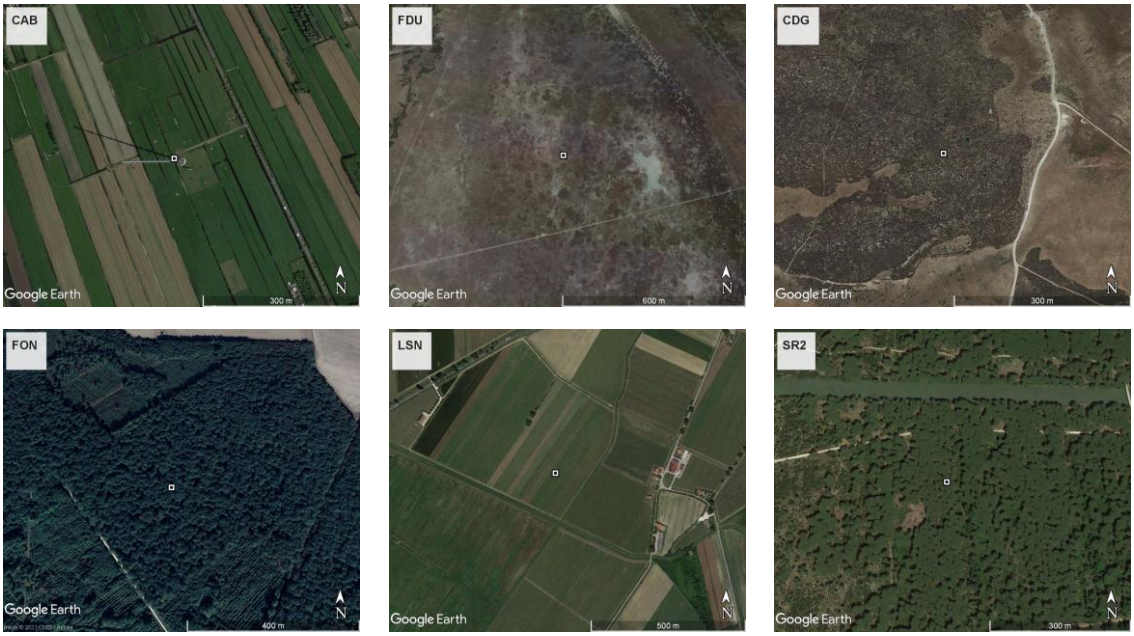


Fig. 2 High resolution images for the nine sites from Google Earth.

3. Evaluation method

3.1. Temperature-based validation

In this study, the Temperature-based validation strategy was adopted for the LST evaluation (Hulley et al. 2021). The different LST products were compared with *in-situ* measurements for the period between August 1, 2018 and December 31, 2021 over Europe and Africa. To mitigate cloud contamination, only pixels surrounded by 15 × 15 (approximately 1 km × 1 km) cloud-free pixels were considered for further evaluation. This is required for a fair evaluation of the ECOSTRESS LST since the ECOSTRESS cloud mask relies only on TIR bands. Meanwhile, the "3σ-Hampel identifier" was adopted to remove the outliers caused by possible cloud contamination or other radiance-related issue (Duan et al. 2019). The standard deviation used in the method is calculated as follows

$$S = 1.4826 \times \text{median}(|x_i - x_m|) \quad (6)$$

where S is the robust standard deviation, x_i is the difference between the LST retrieval and *in-situ* measurement, x_m is the median of the difference. LST retrievals with LST differences below x_m-$3S$ or above x_m+$3S$ were regarded as outliers and excluded from the evaluation. To

ensure the spatial representativeness of the *in-situ* measurements, only match-ups with a standard deviation <1 K within a 3 × 3 window centered on the sites were retained.

Three indices were used to quantify the performance of these LST products based on the recommendation by the Committee on Earth Observation Satellites (CEOS) Working Group on Calibration and Validation - Land Product Validation (LPV) Subgroup (Guillevic et al. 2018). The total uncertainty is estimated via root-mean-square error (RMSE) as follows:

$$RMSE = \sqrt{\frac{\sum(LST_{sat} - LST_{insitu})^2}{N}} \quad (7)$$

where LST_{sat} and LST_{insitu} are the LST products and *in-situ* measurements, respectively, N is the sample number. The accuracy is estimated via bias μ as follows:

$$\mu = \text{median}\left(LST_{sat,i} - LST_{insitu,i}\right). \quad (8)$$

The median is used in lieu of the mean to avoid the impacts of outliers in statistics. Similarly, the median of the absolute residual is calculated as an estimate of the precision σ:

$$\sigma = 1.4826 \times \text{median}\left(\left|\left(LST_{sat,i} - LST_{insitu,i}\right) - \mu\right|\right). \quad (9)$$

3.2. Cross-satellite validation

The Landsat and ASTER-Terra have sun-synchronous orbits. This feature leads to a fixed overpass time, which is around 10 a.m. for Landsat and 10: 30 a.m./p.m. for ASTER. Due to the asynchronous orbits of ISS, the overpass time of ECOSTRESS varies from day to day. It is therefore challenging to obtain Landsat and ASTER LST concurrent with ECOSTRESS LST. Here, Landsat and ASTER LST retrievals were evaluated for the same period at the 9 selected ground sites, and the same accuracy indicators were used as mentioned in Eqs. 7−9.

4. Results and analysis

4.1. Evaluation results using in-situ measurements

Fig. 3 shows the evaluation results of ECOSTRESS LST at the 9 sites. The sample numbers are above 15 at most sites except for the 2 GCU sites due to the sparse coverage of ECOSTRESS over the Iberian Peninsula. The RMSE of the three ECOSTRESS LST are below

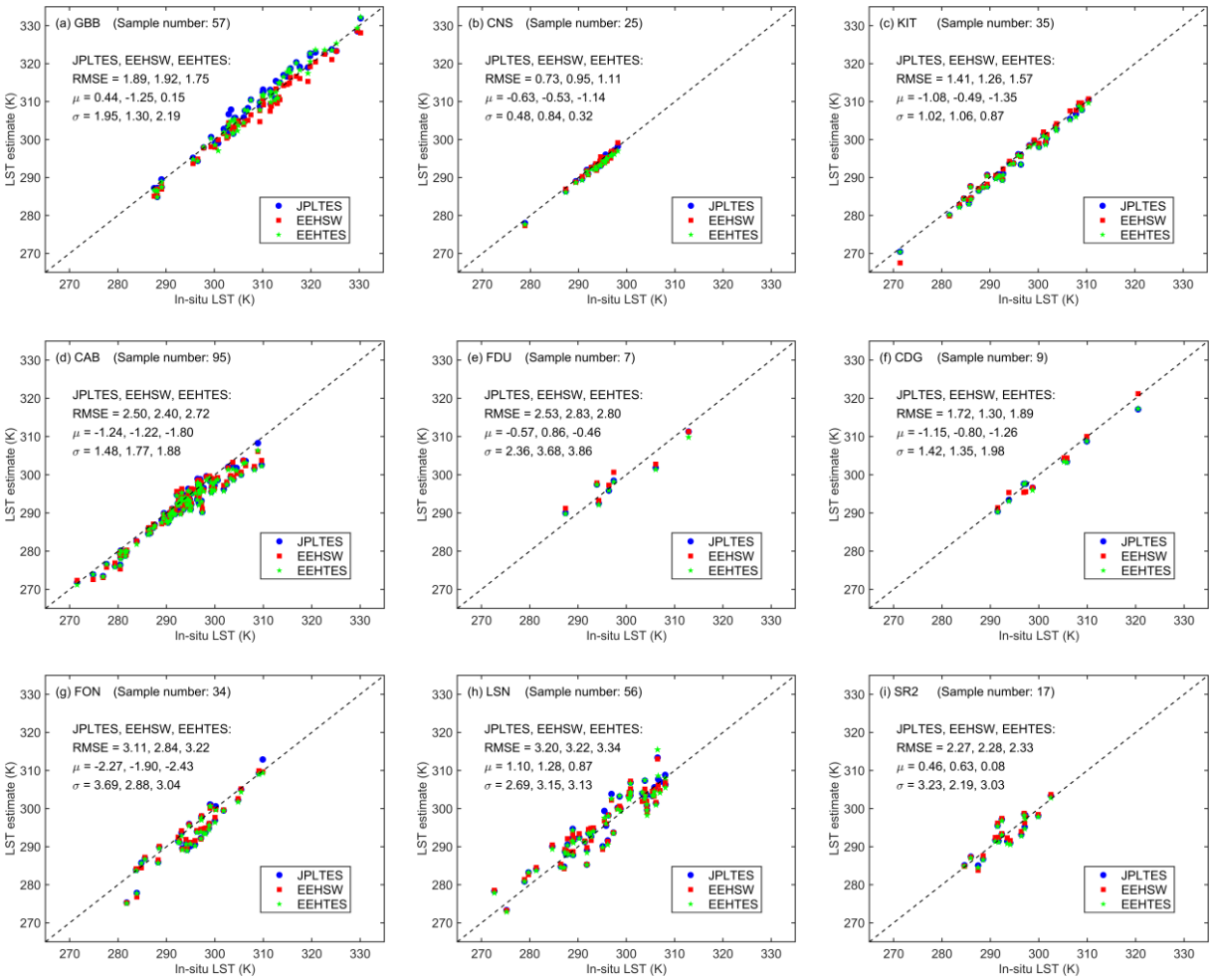
3 K and μ (absolute value) are within 2 K at most sites except for the two ICOS sites FON and LSN. The performances of LST retrievals at GBB, CNS, KIT and CGD are better as compared to the other sites. The RMSE are below 2 K and μ (absolute value) are around 1 K at these 4 sites.

For the three spatially homogeneous KIT and Copernicus sites, the LST products are in good agreement with *in-situ* LST. At GBB and KIT, the RMSEs are around 1.5 K and μ (absolute value) are around 1 K. At CNS, the RMSE and μ (absolute value) are around 1 K. A cold bias exists at all the sites for the three products, except for JPLTES and EEHTES LST at site GBB where LST is generally >295 K. The JPLTES and EEHTES LST are in close agreement at the 3 sites. At the desert site GBB with very dry atmosphere, the EEHTES LST has a slightly better performance (with the RMSE and μ 0.14 and 0.29 K lower, respectively) as compared to the JPLTES LST. At the water site CNS, the difference between JPLTES and EEHTES is more notable, with RMSE and μ of JPLTES approximately 0.4 K lower. For the forest site KIT, the JPLTES LST has a slightly lower RMSE and μ (0.16 and 0.27 K, respectively) as compared to the EEHTES LST. In terms of the EEHSW LST, it has a similar RMSE to the two TES LST at all the three KIT and Copernicus sites. However, μ is found >0.8 K higher at GBB. At the site KIT, the RMSE and μ of EEHSW LST are approximately 0.2 and 0.6 K lower than for the two TES LST, respectively. The better performance of the EEHSW LST at the forest site KIT is probably related to the low spectral contrast of emissivity for dense vegetation, which leads to a degraded performance of the TES algorithm (Gillespie et al. 1998).

At the BSRN site CAB, the RMSE of the JPLTES and EEHSW are close (~2.5 K). The RMSE of the EEHTES LST is ~0.2 K higher than the other two. All the three LST have a negative bias. μ (absolute value) of the EEHTES LST is ~0.5 K higher than the other two, and σ of the JPLTES LST is approximately 1.5 K, which is ~0.3 K lower than those of the EEHSW and EEHTES LST.

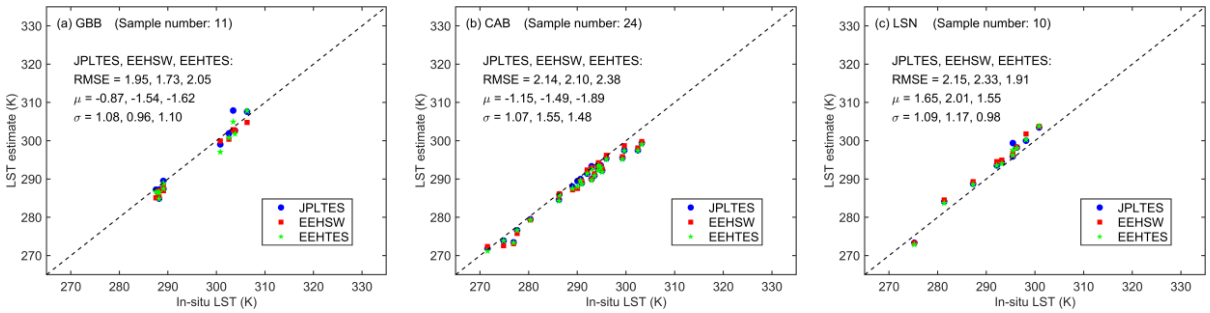
At the two GCU sites, the RMSE of the three ECOSTRESS LST at CDG (<2 K) are substantially lower than at FDU (>2.5 K). This is probably due to the seasonal change of landscape (dry and wet periods) and consequent variant spatial heterogeneity at the marshland site FDU. μ (absolute value) of the JPLTES and EEHTES LST at FDU (<1 K) are lower than at CDG (~1.2 K) while μ (absolute value) of the EEHSW LST (~0.8 K) are similar at both sites. The JPLTES and EEHTES LST are close with respect to RMSE and μ at the two sites, with differences <0.3 K. However, σ of the EEHTES LST is considerably higher than that of the JPLTES LST, especially at FDU where the difference exceeds 1 K. The RMSE of EEHSW LST is similar to those of the two TES LST at FDU (with differences within 0.3 K), but >0.4 K lower as compared to the two TES LST at CDG despite of the sparsely vegetated surface. μ (absolute value) of the EEHSW LST is ~0.3 K higher at FDU but ~0.4 K lower at CDG than the two TES LST.

At the three vegetated ICOS sites, the performances of three ECOSTRESS LST are similar. The RMSE is generally above 3 K at FON and LSN and between 2 and 2.5 K at SR2. μ is around -2 and 1 K at FON and LSN, respectively and is below 1 K at SR2. σ is above 2.5 K in most cases. As compared to the results for the other six sites, LST accuracies at the three ICOS sites are degraded, especially in terms of RMSE. This is probably due to the 30 min sampling interval at the ICOS sites, which is too infrequent to capture the substantial temporal variation in LST at daytime. The two TES LST are close to each other over the ICOS sites, with a difference in RMSE of ~0.1. The EEHSW LST has a ~0.3 K lower RMSE and μ at the forest site FON, which is similar to the case at the forest site KIT. At the sites LSN and SR2, the EEHSW LST is close to the two TES LST with a difference in RMSE within 0.15 K.



370 Fig. 3 ECOSTRESS LST against *in-situ* measurements (K) at the nine ground sites. Daytime and night-time
371 retrievals were included.

372 A comparison of the night-time LST retrievals with *in-situ* measurements reveals no
373 significant improvements in RMSE (Fig. 4) at GBB due to the pronounced spatial homogeneity
374 at the desert site. However, the magnitude of μ is higher during night-time at GBB. At the
375 grassland site CAB, RMSE is approximately 0.3 K lower at night while the difference in μ is
376 negligible between all day and night-time retrievals. At LSN, the RMSE at night is ~1 K lower
377 than for all day, which could be because the temporal variation of LST is minor within the 30
378 min sampling interval and the spatial heterogeneity decreases at night. Similar to GBB and
379 CNS, μ also increases at LSN at night. σ is lower at night in all cases, which is probably due to
380 the narrow range of LST.

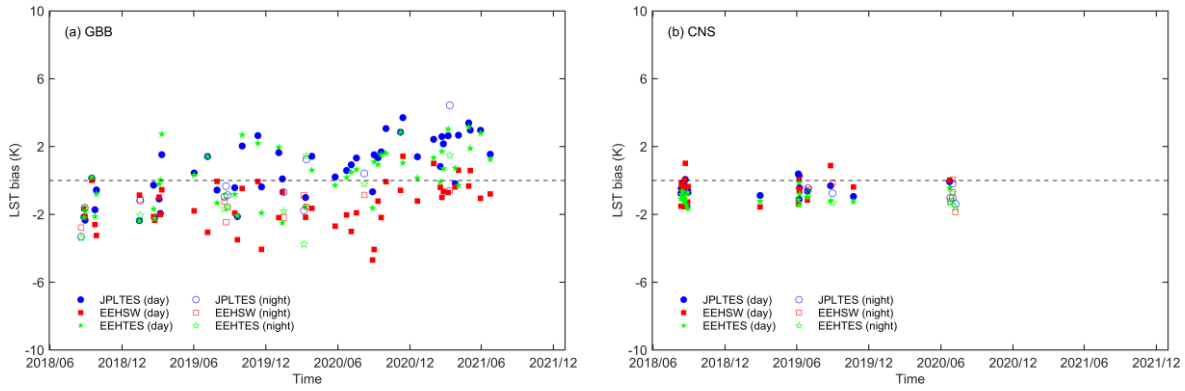


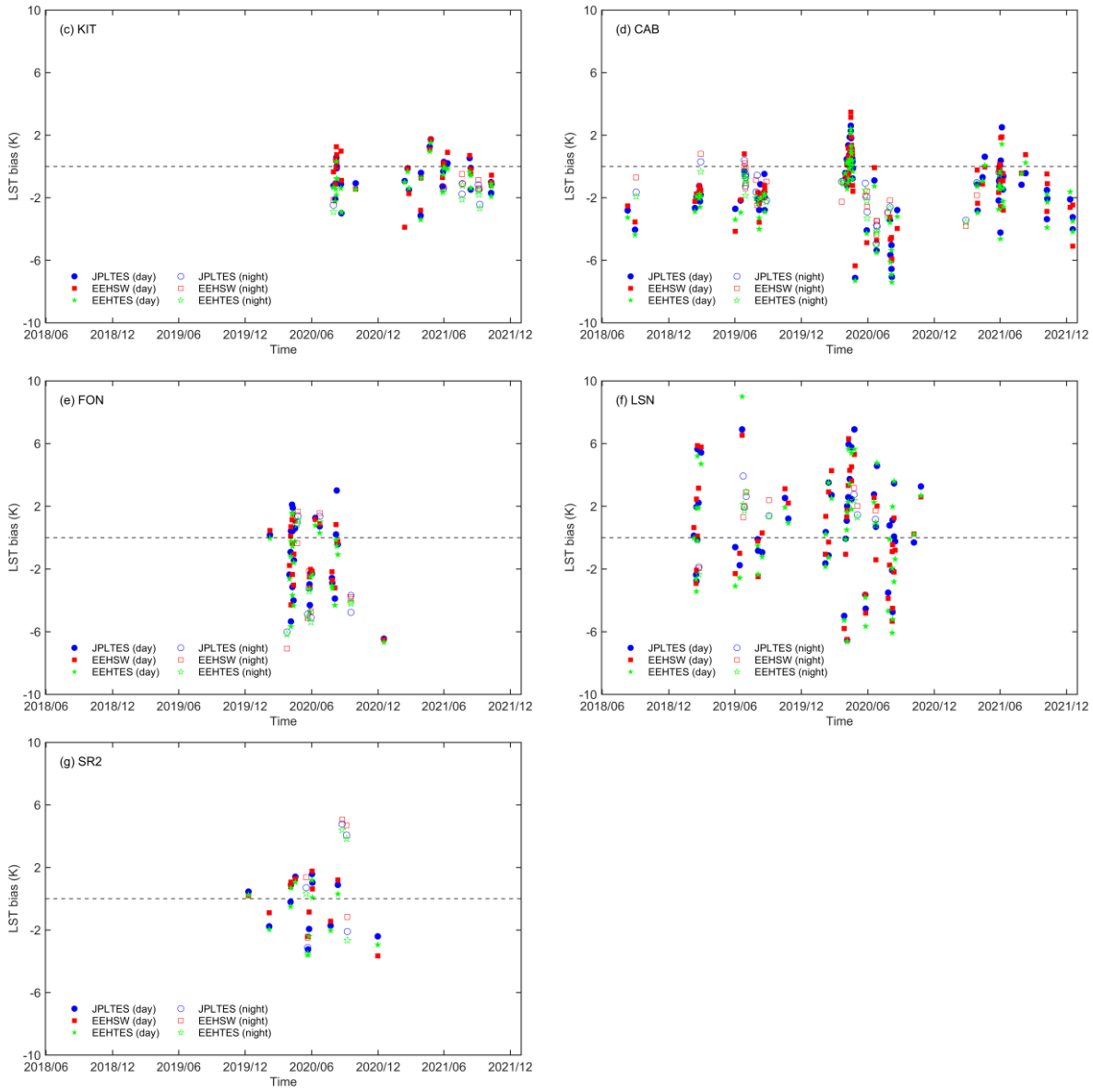
381 Fig. 4 Night-time ECOSTRESS LST against in-situ measurements (K). Only sites with at least 10 match-
382 ups are shown.

383 The LST bias time series (Fig. 5) shows stably distributed bias (±2 K) at the three KIT and
384 Copernicus sites, which is consistent with the good accuracy shown in Fig. 3. At GBB, most
385 samples for EEHSW are distinctively below 0, which is also reflected by the negative μ (-1.25
386 K). The variation ranges of samples for the day and night are similar although the sample
387 number during the night-time is much lower.

388 For CAB, most samples have negative biases, which agrees with the negative μ in Fig. 3.
389 The biases in 2020 are obviously larger, and vary between -8 and 4 K. In the other years, the
390 biases generally range between -6 and 2 K.

391 For the three ICOS sites, the samples are more scattered (varying between ±6 K) as
392 compared to the KIT sites, which is in line with the results in Fig. 3. At LSN, most samples at
393 night are distributed between 0 and 3 K, which leads to a lower RMSE at night than in the day.





394 Fig. 5 Time series of LST bias (ECOSTRESS LST minus *in-situ* measurements) at the ground sites. Sites
395 FDU and CDG are omitted due to the small number of samples.

396 Overall, the three ECOSTRESS LST products agree well with each other, with an RMSE
397 around 2 K and a difference of RMSE within 0.2 K (Fig. 6 and Table 2). The EEHSW LST has
398 the lowest RMSE and σ while the JPLTES LST has the lowest μ (absolute value). μ is within 1
399 K for the JPLTES but slightly above 1 K for the EEHSW and EEHTES LST. A cold bias exists
400 for all the three LST, which is more pronounced for LST below 295 K. This agrees with the
401 finding by Hulley et al. (2021) and relates to the issue in radiometric calibration of the
402 ECOSTRESS radiance data, which is caused by the increased temperature of the cold

blackbody and the increased noise in the cold blackbody. In the future reprocessing of ECOSTRESS data, this cold bias is expected to be mitigated by adjusting the calibration coefficients of the radiance data.

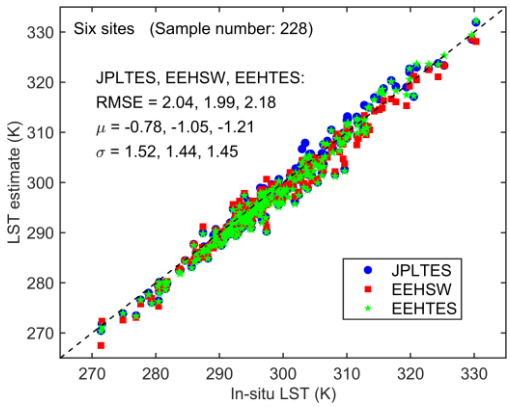


Fig. 6 ECOSTRESS LST retrievals vs. *in-situ* LST at the 6 sites GBB, CNS, KIT, CAB, FUD and CDG. The three ICOS sites were excluded due to their poor performance caused by the 30 min sampling frequency.

Table 2 Statistics for three ECOSTRESS LST products at the 9 sites. All results are in K and the last two rows show summary results for each product.

Sites	JPLTES			EEHSW			EEHTES			Sample
	RMSE	μ	σ	RMSE	μ	σ	RMSE	μ	σ	No.
GBB	1.89	0.44	1.95	1.92	-1.25	1.30	1.75	0.15	2.19	57
CNS	0.73	-0.63	0.48	0.95	-0.53	0.84	1.11	-1.14	0.32	25
KIT	1.41	-1.08	1.02	1.26	-0.49	1.06	1.57	-1.35	0.87	35
CAB	2.50	-1.24	1.48	2.40	-1.22	1.77	2.72	-1.80	1.88	95
FDU	2.53	-0.57	2.36	2.83	0.86	3.68	2.80	-0.46	3.86	7
CDG	1.72	-1.15	1.42	1.30	-0.80	1.35	1.89	-1.26	1.98	9
FON	3.11	-2.27	3.69	2.84	-1.90	2.88	3.22	-2.43	3.04	34
LSN	3.20	1.10	2.69	3.22	1.28	3.15	3.34	0.87	3.13	56
SR2	2.27	0.46	3.23	2.28	0.63	2.19	2.33	0.08	3.03	17
6 sites*	2.04	-0.78	1.52	1.99	-1.05	1.44	2.18	-1.21	1.45	228
All	2.41	-0.67	2.05	2.35	-0.87	1.89	2.54	-1.12	2.00	335

*6 sites include GBB, CNS, KIT, CAB, FDU and CDG.

4.2. Inter-comparison with Landsat and ASTER LST retrievals

4.2.1. Inter-comparison with Landsat LST

The match-ups for all 9 sites are at least 15. The two GCU sites are covered considerably better by Landsat than by ECOSTRESS. Overall, the RMSE is within 3 K and μ within 1.5 K at most sites (except for FDU, FON and LSN).

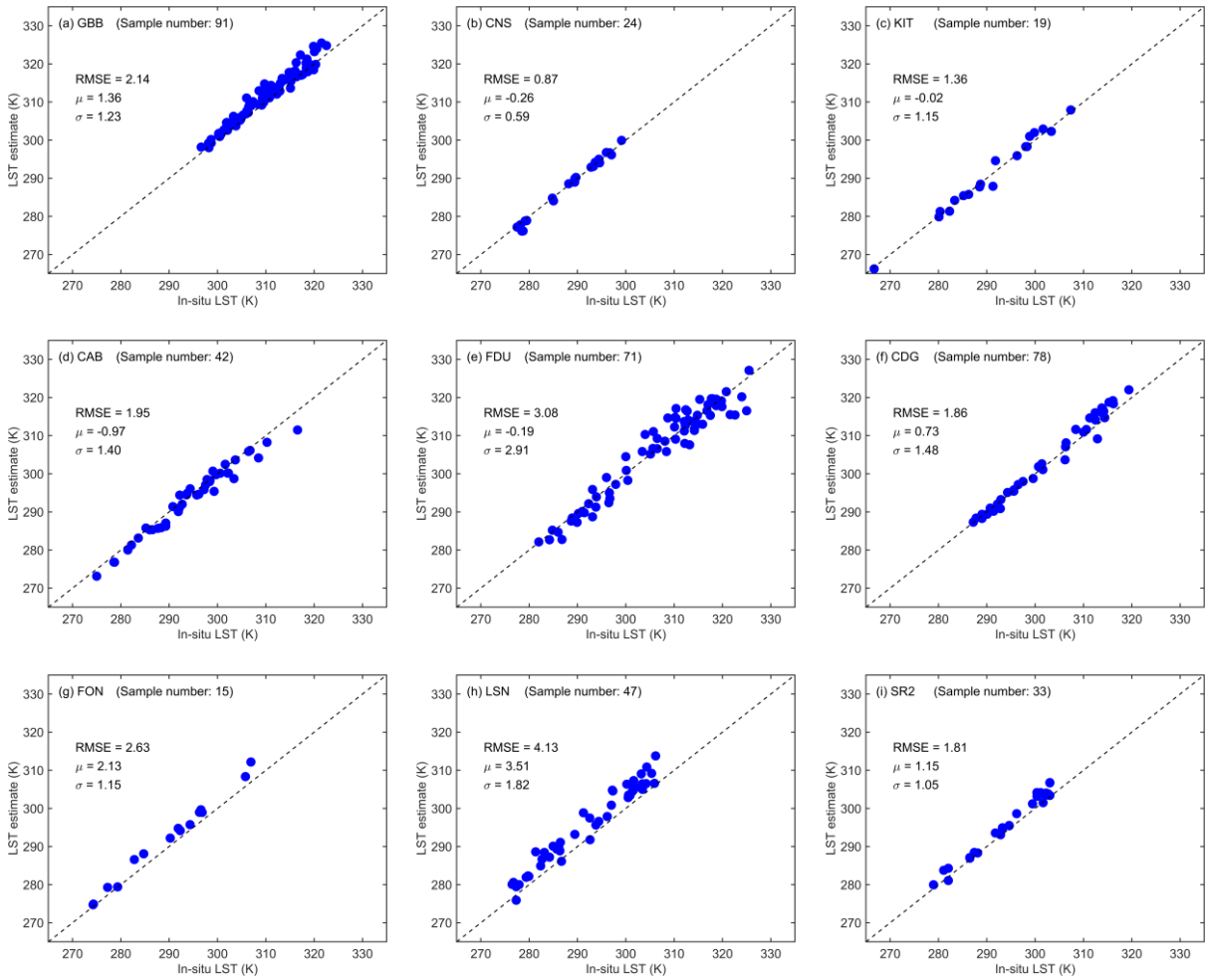


Fig. 7 Landsat LST against in-situ measurements (K) at the ground sites.

The Landsat LST retrievals show a comparably good accuracy at the 3 KIT and Copernicus sites. At GBB, the RMSE is slightly higher (~0.3 K) than the three ECOSTRESS LST. μ for Landsat LST (absolute value) is close to that for the EEHSW LST that similarly uses the ASTER GED data for emissivity retrieval but higher (>0.8 K) than for the TES LST, which is in line with the higher μ of the EEHSW LST as compared to the JPLTES and EEHTES LST (Table 2). At the water site CNS, the performance of Landsat LST is close to those of the three ECOSTRESS LST, with an RMSE ~1 K. At the forest site KIT, the RMSE (~1.5 K) is also close to those of the three ECOSTRESS LST. μ (absolute value) of Landsat LST is below 0.5 K, which is similar to the EEHSW LST, but more than 1 K lower than for the two TES LST.

At CAB, the RMSE of Landsat LST is ~0.5 K lower than the three ECOSTRESS LST. μ is ~0.2 K lower than for the JPLTES and EEHSW LST and ~0.8 K lower than for the EEHTES LST.

As found for the ECOSTRESS LST, at the two GCU sites the performance of the Landsat LST at CDG is much better than at FDU. The RMSE of the Landsat LST is above 3 K at FDU, which is higher (0.25−0.55 K) than for the three ECOSTRESS LST. μ (absolute value) of Landsat LST approaches 0 K at FDU, which is lower than for the three ECOSTRESS LST. At CDG, the RMSE of the Landsat LST is close to those of the two ECOSTRESS TES LST (~1.8 K) but ~0.5 K higher than that of the EEHSW LST. μ (absolute value) of the Landsat LST is close to that of the EEHSW LST (~0.8 K) and ~0.4 K lower than those of the two ECOSTRESS TES LST.

At the ICOS sites, the accuracy of Landsat LST is also lower than at the other sites. At FON, the RMSE of Landsat LST is similar to that of EEHSW (~2.5 K), which is around 0.6 K lower than for the two TES LST. At LSN, a significant overestimation exists in Landsat LST, which leads to an RMSE and μ of Landsat LST around 1 and 2 K higher than for the three ECOSTRESS LST, respectively. At SR2, the RMSE of Landsat LST is ~0.4 K lower than for the three ECOSTRESS LST. However, μ is more than 0.5 K higher than for the ECOSTRESS LST.

The validation results of Landsat LST for the six sites (excluding ICOS sites) are very similar to those of ECOSTRESS LST (Fig. 8). The RMSE is 2.20 K, which is ~0.2 K higher than for the three ECOSTRESS LST. μ (absolute value) of Landsat LST is close to that of JPLTES LST (~0.8 K) and slightly lower than for the EEHSW and EEHTES LST, but with a different sign. σ of Landsat LST is similar to those of the three ECOSTRESS LST (~1.5 K), although it is generally lower than for the ECOSTRESS LST at individual sites.

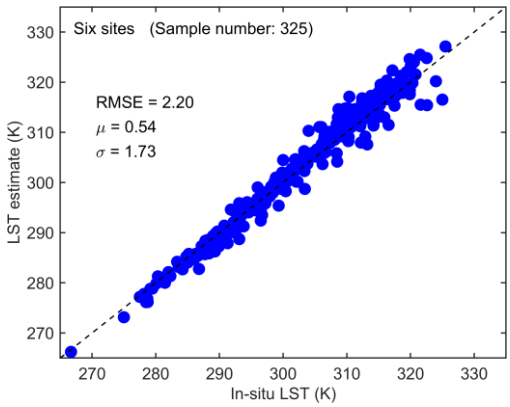
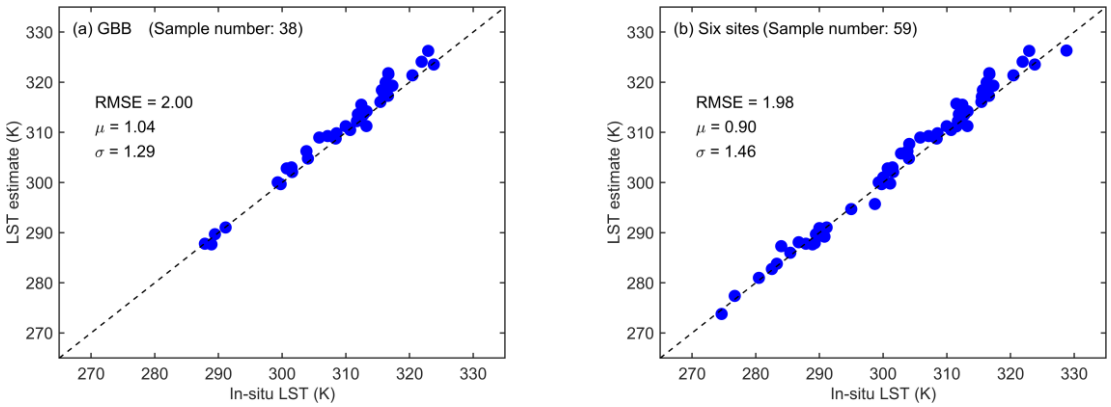


Fig. 8 Landsat LST retrievals vs. *in-situ* LST at the 6 sites GBB, CNS, KIT, CAB, FUD and CDG. The three ICOS sites were excluded due to their poor performance caused by the 30 min sampling frequency.

4.2.2. Inter-comparison with ASTER LST

The ASTER LST have sparse coverages at most of the nine sites except for GBB, which provides 38 matchups (Fig. 9). At GBB, the RMSE (2 K) of ASTER LST is slightly higher (~0.1 K) than for the three ECOSTRESS LST and slightly lower (~0.1 K) than for the Landsat LST. μ (absolute value) of ASTER LST is close to those of ECOSTRESS SW and Landsat LST (~1 K), which use the ASTER emissivity, but higher (>0.5 K) than for the two ECOSTRESS TES LST. σ of ASTER LST is also similar to that of the ECOSTRESS SW and Landsat LST (~1.3 K) and lower (~0.6 K) than for the two ECOSTRESS TES LST.

Overall, the RMSE of the ASTER LST is 1.98 K, which is similar to those for the ECOSTRESS LST and ~0.2 K lower than for the Landsat LST (Tables 2 and 3). μ (absolute value) is close to the ECOSTRESS and Landsat LST, which is approximately 1 K. σ of the ASTER LST is similar to those of the ECOSTRESS LST (~1.5 K), but lower (~0.3 K) than for the Landsat LST.



467 Fig. 9 ASTER LST against *in-situ* measurements (K) at (a) GBB and (b) at the 6 sites GBB, CNS, KIT, CAB,
468 FUD and CDG.

469 Table 3 Statistics results for Landsat and ASTER LST products at the 9 sites. The average RMSE, μ and σ
470 for ASTER LST are in K for all available samples with a required minimum sample number of 5.

Sites	Landsat				ASTER			
	RMSE	μ	σ	Sample No.	RMSE	μ	σ	Sample No.
GBB	2.14	1.36	1.23	91	2.00	1.04	1.29	38
CNS	0.87	-0.26	0.59	24	-	-	-	2
KIT	1.36	-0.02	1.15	19	-	-	-	2
CAB	1.95	-0.97	1.40	42	-	-	-	5
FDU	3.08	-0.19	2.91	71	-	-	-	6
CDG	1.86	0.73	1.48	78	-	-	-	6
FON	2.63	2.13	1.15	15	-	-	-	0
LSN	4.13	3.51	1.82	47	-	-	-	7
SR2	1.81	1.15	1.05	33	-	-	-	5
6 sites*	2.20	0.54	1.73	325	1.98	0.90	1.46	59
All	2.48	0.85	1.94	420	2.06	0.90	1.62	71

471 Among the five LST retrievals (Fig. 10), the JPLTES, EEHSW and ASTER have similar
472 RMSE (approximately 2 K) and are ~0.2 K lower than for the EEHTES and Landsat LST. The
473 Landsat LST has the lowest μ (absolute value) but the highest RMSE and σ. The JPLTES,
474 EEHSW, EEHTES and ASTER have similar σ (~1.5 K) and are ~0.2 K lower than for the
475 Landsat LST. In general, the five high spatial resolution LST products show a high level of
476 consistency, with an RMSE around 2 K, μ around 1 K and σ around 1.5 K.

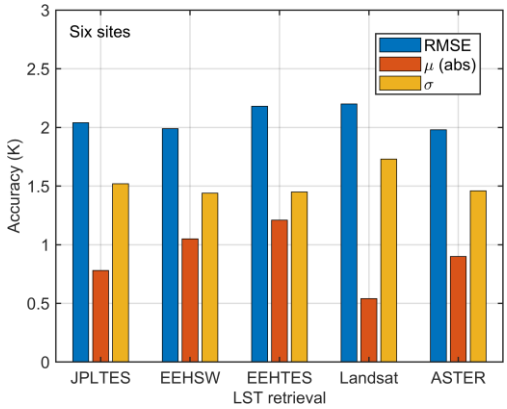


Fig. 10 Comparison among the five LST products in terms of RMSE, μ and σ for the six sites GBB, CNS, KIT, CAB, FUD and CDG. The absolute value of μ is shown and μ is negative for the JPLTES, EEHSW and EEHTES ECOSTRESS LST.

5. Discussion

5.1. Dependence of LST retrieval uncertainty on water vapor content and vegetation condition

Atmospheric and vegetation conditions are two important factors for LST retrieval. Here, the dependence of LST bias on precipitable water vapor (PWV) content and normalized differential vegetation index (NDVI) was investigated (Fig. 11). No clear trend is evident in the scatterplot between LST bias and PWV/NDVI. For the RMSE in different PWV intervals, an increase is noted (>0.5 K) in the 3−5 cm interval. The RMSE for the JPLTES and EEHTES LST are almost identical when PWV is below 2 cm. Whereas, the RMSE of the EEHTES LST is clearly higher than that of the JPLTES LST when PWV is above 2 cm. This agrees with the fact that the difference between the JPLTES and EEHTES LST mainly stems from the different atmospheric profiles used in atmospheric correction for these two LST products. The EEHSW LST has the lowest RMSE in most cases except when PWV is below 1 cm. This indicates that the LST retrieval accuracy from TES is more sensitive to PWV as compared to that from SW.

For the RMSE in different NDVI intervals, an increase for NDVI greater than 0.4 is evident. We infer that this is attributed to an increase in uncertainties of emissivity retrievals from the TES algorithm with vegetation coverage, which is also propagated via the ASTER GED data to the EEHSW LST. The RMSE for the NDVI interval between 0.4 and 0.6 (>3 K) is notably

higher than for the other intervals. This is probably due to the increased spatial heterogeneity generally observed for intermediate vegetation coverage, which can reduce the spatial representativeness of the ground measurements.

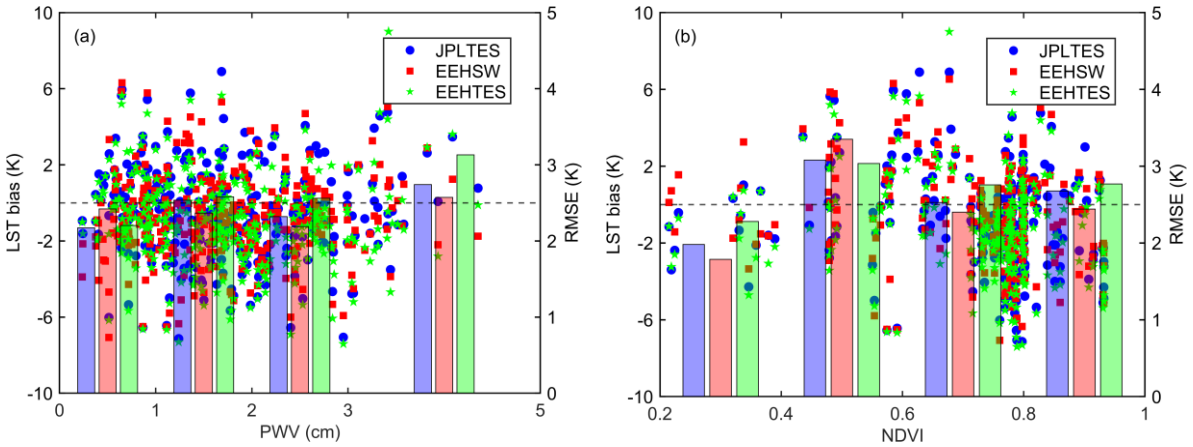


Fig. 11 LST bias (LST retrieval minus *in-situ* measurements) versus (a) PWV and (b) NDVI at all the sites. PWV was extracted from the ERA5 data, same as that used in the EEHSW algorithm. NDVI was obtained from the MOD13Q1 data after outlier removal and time series smoothing. The scatterplots show the LST biases for all the samples across different PWV and NDVI values. The bar plots show the RMSE of LST retrievals for each PWV and NDVI interval.

5.2. Comparison among different LST estimation algorithms

Different algorithms were used to retrieve LST from a host of satellite data, including the single-channel algorithm for Landsat, the TES algorithm for ASTER, and the TES and SW algorithms for ECOSTRESS. However, the overall accuracies for these LST were close. The EEHSW LST had the best correlation with the Landsat LST with regard to the three indicators among the three ECOSTRESS LST.

For the Landsat, EEHSW and ASTER LST, which are generated with ASTER emissivities but different retrieval algorithms, they performed very consistently at GBB (see Tables 2 and 3), with a higher μ (absolute value) and a lower σ as compared to the two ECOSTRESS TES LST. This confirms that emissivity retrieval plays a critical role in LST estimation and that LST retrievals using consistent emissivities could achieve similar accuracies when the atmosphere is dry.

5.3. Challenges of Temperature-based validation

Spatial and temporal representativeness of ground measurements are among the major challenges existing for the Temperature-based validation. The FOVs of ground measurements are normally much smaller than the satellite pixels, even though as compared to the high spatial resolution LST data (Guillevic et al. 2014). Therefore, the area surrounding ground sites should be homogeneous at the pixel scale to ensure the spatial representativeness of the ground measurements. This explains the better performance of the LST retrievals at the KIT and Copernicus sites. The three KIT and Copernicus sites have been specifically designed to evaluate satellite LST products (Göttsche et al. 2013; Göttsche et al. 2016). The homogeneous landscapes surrounding the three sites (i.e., desert, water, and dense forest), choice of radiometers and regular maintenance ensure the accuracy and credibility of the validation results.

Sampling frequency is also an important factor to consider in LST validation. For KIT, Copernicus, BSRN and GCU sites, the sampling frequency is within 5 min, which is appropriate for validating instantaneous LST retrievals. However, the sampling frequency of the ICOS sites is 30 min, which is insufficient for LST validation, especially for fast changing LST in the morning. This is reflected in the generally larger uncertainties of the validation results at the ICOS sites. Therefore, the results at the ICOS sites are not included in the overall statistics.

Different sensors are used at the ground sites. At the KIT, Copernicus and GCU sites that were designed for LST validation, the precision radiometer KT15.85 IIP and the Apogee broadband radiometer are utilized. At the BSRN and ICOS sites that were set up for studying surface radiation budget and carbon and energy fluxes, respectively, pyrgeometers are used to measure the upwelling and downwelling broadband hemispheric longwave radiations. However, the directional temperature estimated from the radiance measurement is closer to

LST retrieved from satellite observations, which is further amplified over heterogeneous surfaces (Trigo et al. 2021; Yang et al. 2020). This partially explains why the RMSE at the KIT and GCU sites were generally lower.

6. Conclusions and perspectives

Launched to the ISS in 2018, ECOSTRESS serves as a bridge between the Landsat and ASTER missions that provide LST products at high spatial but coarse temporal resolutions and the next generation of TIR missions with high spatiotemporal resolutions (i.e., TRISHNA, LSTM and SBG). In this study, we evaluated three different ECOSTRESS LST products, i.e., the JPLTES, EEHSW and EEHTES LST, over Europe and Africa by comparing them with *in-situ* measurements and the corresponding LST products from Landsat and ASTER at 9 ground sites. To the best of our knowledge, this is the first study that demonstrates a comprehensive comparison among the state-of-the-art high spatial resolution LST data.

The RMSE of the three ECOSTRESS LST are below 3 K and μ (absolute value) are within 2 K at all sites but the two ICOS sites FON and LSN, which could be related to the coarse sampling frequency (30 min) of the ICOS sites and physical differences between hemispherically effective temperature and satellite directional temperature. The performances of LST retrievals at the homogeneous sites GBB, CNS, KIT and CDG are better than at the other sites, which also corresponds to the high sampling frequencies of in-situ measurements and choice of sensor. The RMSE are below 2 K and μ (absolute value) are around 1 K at these 4 sites. Overall, the three ECOSTRESS LST perform consistently, with an RMSE around 2 K and differences in RMSE within 0.2 K. A cold bias around 1 K exists for all three LST. This is related to the radiometric calibration of the ECOSTRESS Level 1 radiance data in Collection 1.

The validation results of Landsat LST are close to those of ECOSTRESS LST, with an overall RMSE of 2.20 K and differences from the three ECOSTRESS LST within 0.2 K. The

EEHSW LST has the best consistency with the Landsat LST, probably due to the identical emissivity retrieval process. The RMSE of the ASTER LST is 1.98 K. μ (absolute value) is close to the ECOSTRESS and Landsat LST, which is approximately 1 K. In general, the five high spatial resolution LST retrievals show a high level of consistency, with an RMSE around 2 K, bias around 1 K, σ around 1.5 K.

In the future, the ECOSTRESS radiance data will be reprocessed in Collection 2 to correct for the cold bias of the radiance calibration. This will further improve the LST accuracy. The ECOSTRESS LST retrievals, together with observations from the future high spatiotemporal resolution TIR missions, will provide unprecedented opportunities for studies in land surface monitoring and terrestrial ecosystem functioning.

Acknowledgement

The authors wish to extend their gratitude to all the scientists involved in the ECOSTRESS mission and ground measurement collection. The study was conducted under the European ECOSTRESS Hub project (EEH, Contract No. 4000129873/20/I-NS), funded by the European Space Agency (ESA).

Appendix

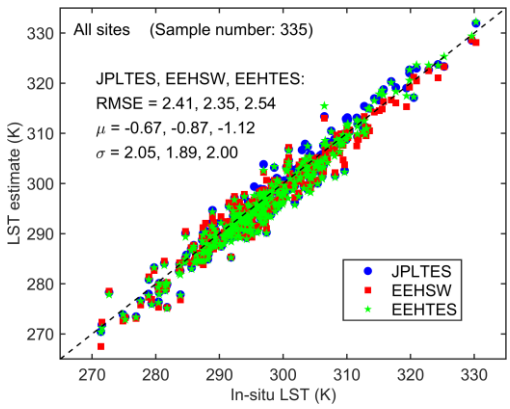
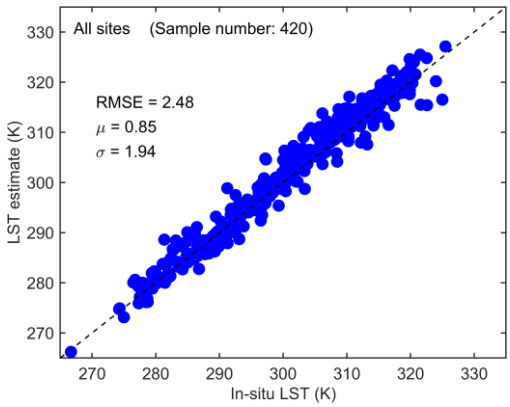
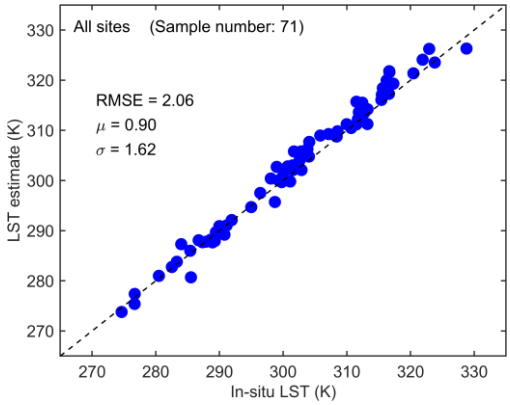


Fig. A1 ECOSTRESS LST retrievals vs. *in-situ* LST by assembling the samples at all the sites.



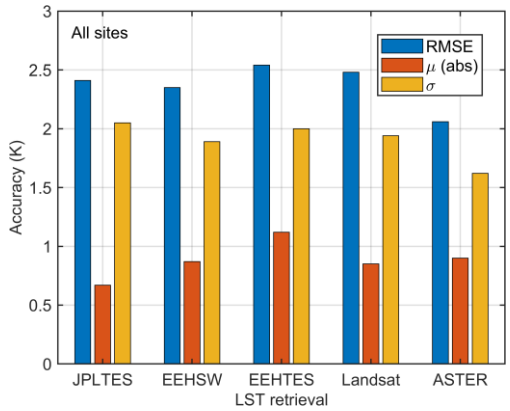
586

587 Fig. A2 Landsat LST retrievals vs. *in-situ* LST by assembling the samples at all the sites.



588

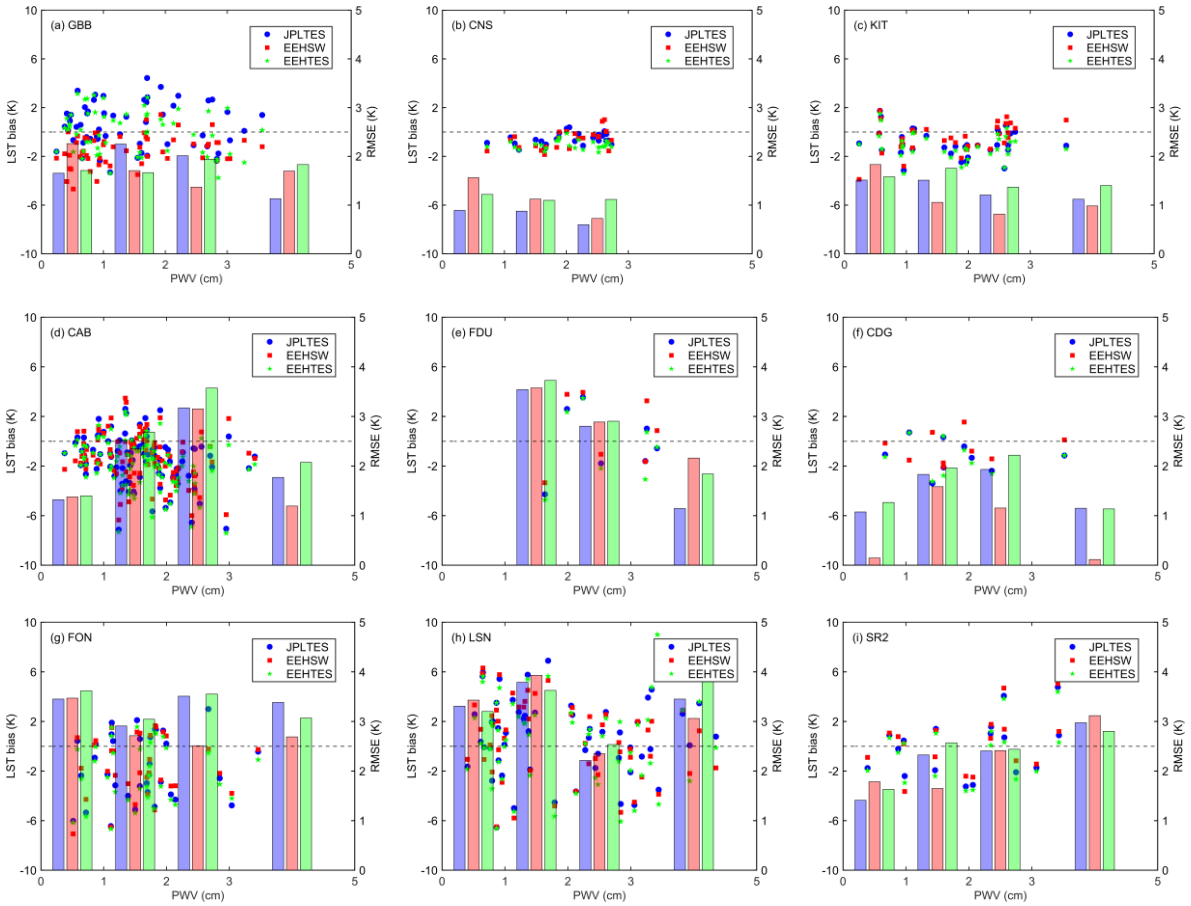
589 Fig. A3 ASTER LST retrievals vs. *in-situ* LST by assembling the samples at all the sites.



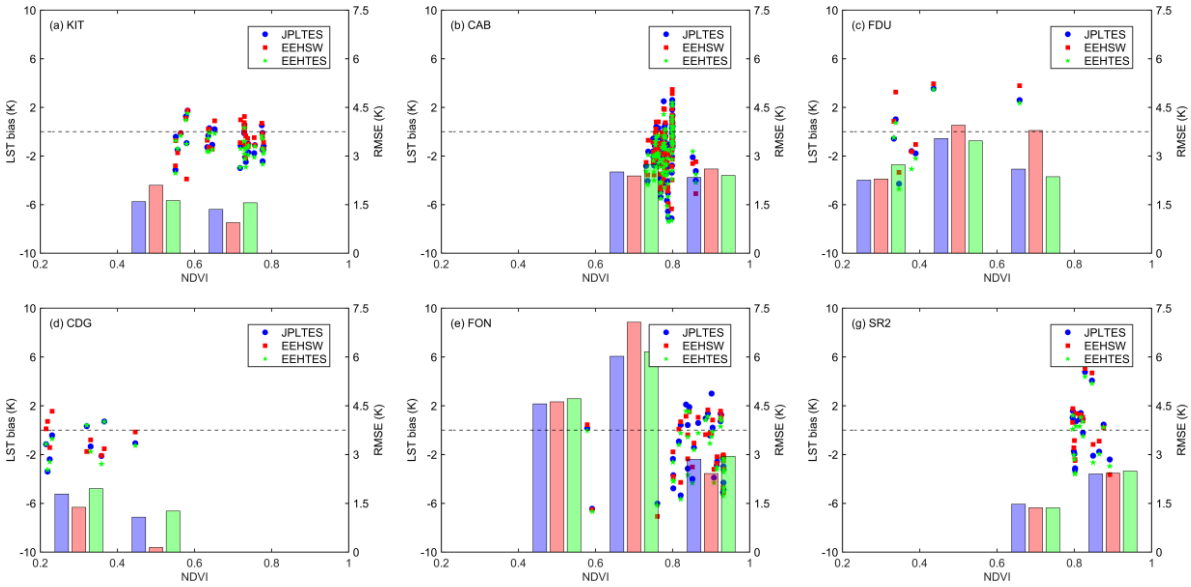
590

591 Fig. A4 Comparison among the five LST products in terms of RMSE, μ and σ for all the nine sites. The

592 absolute value of μ is shown and μ is negative for the JPL, EEHSW and EEHTES ECOSTRESS LST.



593 Fig. A5 LST bias (LST retrieval minus *in-situ* measurements) versus PWV at the 9 sites. PWV was extracted
594 from the ERA5 data, same as that used in the EEHSW algorithm. The scatterplots show individual LST
595 biases for their respective PWV values. The bar plots show the RMSE of LST retrievals for each PWV
596 interval.



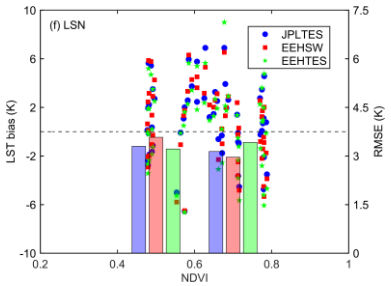


Fig. A6 LST bias (LST retrieval minus *in-situ* measurements) versus NDVI at the selected sites. The desert site GBB and water site CNS were excluded (lack of vegetation). NDVI was obtained from the MOD13Q1 data after outlier removal and time series smoothing. The scatterplots show individual LST biases for their respective NDVI values. The bar plots show the RMSE of LST retrievals for each NDVI interval.

Reference

Abrams, M. (2000). The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform. *International Journal of Remote Sensing, 21*, 847-859

Abrams, M., Tsu, H., Hulley, G., Iwao, K., Pieri, D., Cudahy, T., & Kargel, J. (2015). The advanced spaceborne thermal emission and reflection radiometer (ASTER) after fifteen years: review of global products. *International Journal of Applied Earth Observation and Geoinformation, 38*, 292-301

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 sensing of environment, 122*, 50-65

Anderson, M.C., Yang, Y., Xue, J., Knipper, K.R., Yang, Y., Gao, F., Hain, C.R., Kustas, W.P., Cawse-Nicholson, K., & Hulley, G. (2021). Interoperability of ECOSTRESS and Landsat for mapping evapotranspiration time series at sub-field scales. *Remote sensing of environment, 252*, 112189

Anderson, M.C., Zolin, C.A., Sentelhas, P.C., Hain, C.R., Semmens, K., Yilmaz, M.T., Gao, F., Otkin, J.A., & Tetrault, R. (2016). The Evaporative Stress Index as an indicator of

agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote sensing of environment, 174*, 82-99

Bai, Y., Bhattarai, N., Mallick, K., Zhang, S., Hu, T., & Zhang, J. (2022). Thermally derived evapotranspiration from the Surface Temperature Initiated Closure (STIC) model improves cropland GPP estimates under dry conditions. *Remote sensing of environment*, *271*, 112901

Becker, F., & Li, Z.L. (1995). Surface temperature and emissivity at various scales: Definition, measurement and related problems. *Remote Sensing Reviews, 12*, 225-253

Cawse-Nicholson, K., Townsend, P.A., Schimel, D., Assiri, A.M., Blake, P.L., Buongiorno, M.F., Campbell, P., Carmon, N., Casey, K.A., & Correa-Pabón, R.E. (2021). NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. *Remote sensing of environment, 257*, 112349

Chakraborty, T.C., Lee, X., Ermida, S., & Zhan, W. (2021). On the land emissivity assumption and Landsat-derived surface urban heat islands: A global analysis. *Remote sensing of environment, 265*, 112682

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., Denn, F.M., Duprat, T., Fukuda, M., & Grobe, H. (2018). Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017). *Earth System Science Data, 10*, 1491-1501

Duan, S.-B., Li, Z.-L., Li, H., Göttsche, F.-M., Wu, H., Zhao, W., Leng, P., Zhang, X., & Coll, C. (2019). Validation of Collection 6 MODIS land surface temperature product using in situ measurements. *Remote sensing of environment, 225*, 16-29

Ekinzog, E.K., Schlerf, M., Kraft, M., Werner, F., Riedel, A., Rock, G., & Mallick, K. (2022). Revisiting crop water stress index based on potato field experiments in Northern Germany. *Agricultural Water Management, 269*, 107664

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse- Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., & Arain, M.A. (2020). ECOSTRESS: NASA's next

generation mission to measure evapotranspiration from the International Space Station. *Water Resources Research, 56*, e2019WR026058

Freitas, S.C., Trigo, I.F., Bioucas-Dias, J.M., & Gottsche, F.-M. (2009). Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 523-534

Gillespie, A., Rokugawa, S., Matsunaga, T., Cothern, J.S., Hook, S., & Kahle, A.B. (1998). A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. *IEEE Transactions on Geoscience and Remote Sensing, 36*, 1113-1126

Göttsche, F.-M., Olesen, F.-S., & Bork-Unkelbach, A. (2013). Validation of land surface temperature derived from MSG/SEVIRI with in situ measurements at Gobabeb, Namibia. *International Journal of Remote Sensing, 34*, 3069-3083

Göttsche, F.-M., Olesen, F.-S., Trigo, I.F., Bork-Unkelbach, A., & Martin, M.A. (2016). Long term validation of land surface temperature retrieved from MSG/SEVIRI with continuous in-situ measurements in Africa. *Remote Sensing, 8*, 410

Guangmeng, G., & Mei, Z. (2004). Using MODIS land surface temperature to evaluate forest fire risk of northeast China. *IEEE Geoscience and Remote sensing letters, 1*, 98-100

Guillevic, P., Göttsche, F., Nickeson, J., Hulley, G., Ghent, D., Yu, Y., Trigo, I., Hook, S., Sobrino, J.A., & Remedios, J. (2018). Land surface temperature product validation best practice protocol. Version 1.1. *Best Practice for Satellite-Derived Land Product Validation, 60*

Guillevic, P.C., Biard, J.C., Hulley, G.C., Privette, J.L., Hook, S.J., Olioso, A., Göttsche, F.M., Radocinski, R., Román, M.O., & Yu, Y. (2014). Validation of Land Surface Temperature products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements. *Remote sensing of environment, 154*, 19-37

667 Hook, S.J., Cawse-Nicholson, K., Barsi, J., Radocinski, R., Hulley, G.C., Johnson, W.R.,
668 Rivera, G., & Markham, B. (2019). In-flight validation of the ECOSTRESS, Landsats 7 and 8
669 thermal infrared spectral channels using the Lake Tahoe CA/NV and Salton Sea CA automated
670 validation sites. *IEEE Transactions on Geoscience and Remote Sensing, 58*, 1294-1302
671 Hu, T., Du, Y., Cao, B., Li, H., Bian, Z., Sun, D., & Liu, Q. (2016). Estimation of upward
672 longwave radiation from vegetated surfaces considering thermal directionality. *IEEE*
673 *Transactions on Geoscience and Remote Sensing, 54*, 6644-6658
674 Hu, T., Renzullo, L.J., van Dijk, A.I.J.M., He, J., Tian, S., Xu, Z., Zhou, J., Liu, T., & Liu, Q.
675 (2020). Monitoring agricultural drought in Australia using MTSAT-2 land surface temperature
676 retrievals. *Remote sensing of environment, 236*, 111419
677 Hulley, G.C., Göttsche, F.M., Rivera, G., Hook, S.J., Freepartner, R.J., Martin, M.A., Cawse-
678 Nicholson, K., & Johnson, W.R. (2021). Validation and quality assessment of the
679 ECOSTRESS level-2 land surface temperature and emissivity product. *IEEE Transactions on*
680 *Geoscience and Remote Sensing*
681 Hulley, G.C., Hook, S.J., Abbott, E., Malakar, N., Islam, T., & Abrams, M. (2015). The ASTER
682 Global Emissivity Dataset (ASTER GED): Mapping Earth's emissivity at 100 meter spatial
683 scale. *Geophysical Research Letters, 42*, 7966-7976
684 Islam, T., Hulley, G.C., Malakar, N.K., Radocinski, R.G., Guillevic, P.C., & Hook, S.J. (2016).
685 A physics-based algorithm for the simultaneous retrieval of land surface temperature and
686 emissivity from VIIRS thermal infrared data. *IEEE Transactions on Geoscience and Remote*
687 *Sensing, 55*, 563-576
688 Koetz, B., Bastiaanssen, W., Berger, M., Defourney, P., del Bello, U., Drusch, M., Drinkwater,
689 M., Duca, R., Fernandez, V., & Ghent, D. (2019). Agriculture: Land Surface Temperature
690 Monitoring (LSTM) Mission. In, *ESA Living Planet Symposium*. Milan, Italy

691 Lagouarde, J.P., Bhattacharya, B.K., Crebassol, P., Gamet, P., Babu, S.S., Boulet, G., Briottet,
692 X., Buddhiraju, K.M., Cherchali, S., & Dadou, I. (2018). The Indian-French TRISHNA
693 mission: Earth observation in the thermal infrared with high spatio-temporal resolution. In,
694 *2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4078-4081).
695 Valencia, Spain: IEEE
696 Li, H., Sun, D., Yu, Y., Wang, H., Liu, Y., Liu, Q., Du, Y., Wang, H., & Cao, B. (2014).
697 Evaluation of the VIIRS and MODIS LST products in an arid area of Northwest China. *Remote*
698 *sensing of environment, 142*, 111-121
699 Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., & Sobrino, J.A. (2013).
700 Satellite-derived land surface temperature: Current status and perspectives. *Remote sensing of*
701 *environment, 131*, 14-37
702 Liang, S., Wang, K., Zhang, X., & Wild, M. (2010). Review on estimation of land surface
703 radiation and energy budgets from ground measurement, remote sensing and model simulations.
704 *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3*, 225-
705 240
706 Liu, L., & Zhang, Y. (2011). Urban heat island analysis using the Landsat TM data and ASTER
707 data: A case study in Hong Kong. *Remote Sensing, 3*, 1535-1552
708 Malakar, N.K., Hulley, G.C., Hook, S.J., Laraby, K., Cook, M., & Schott, J.R. (2018). An
709 operational land surface temperature product for Landsat thermal data: Methodology and
710 validation. *IEEE Transactions on Geoscience and Remote Sensing, 56*, 5717-5735
711 Mallick, K., Boegh, E., Trebs, I., Alfieri, J.G., Kustas, W.P., Prueger, J.H., Niyogi, D., Das, N.,
712 Drewry, D.T., & Hoffmann, L. (2015). Reintroducing radiometric surface temperature into the
713 P enman- M onteith formulation. *Water Resources Research, 51*, 6214-6243

714 Mallick, K., Jarvis, A.J., Boegh, E., Fisher, J.B., Drewry, D.T., Tu, K.P., Hook, S.J., Hulley,
715 G., Ardö, J., & Beringer, J. (2014). A Surface Temperature Initiated Closure (STIC) for surface
716 energy balance fluxes. *Remote sensing of environment, 141*, 243-261
717 Mallick, K., Toivonen, E., Trebs, I., Boegh, E., Cleverly, J., Eamus, D., Koivusalo, H., Drewry,
718 D., Arndt, S.K., & Griebel, A. (2018). Bridging Thermal Infrared Sensing and Physically-
719 Based Evapotranspiration Modeling: From Theoretical Implementation to Validation Across
720 an Aridity Gradient in Australian Ecosystems. *Water Resources Research, 54*, 3409-3435
721 McMillin, L.M. (1975). Estimation of sea surface temperatures from two infrared window
722 measurements with different absorption. *Journal of geophysical research, 80*, 5113-5117
723 NASA (2005). Exploring our planet for the benefit of society, NASA earth science and
724 applications from space, strategic roadmap. In.
725 http://images.spaceref.com/news/2005/earth_roadmap.pdf
726 NASA (2011). NASA earth science data records programs. In. https://science.nasa.gov/earth-
727 science/earth-science-data/Earth-Science-Data-Records-
728 Programs/#ESDR_uncertainty_analysis
729 Qin, Z., Karnieli, A., & Berliner, P. (2001). A mono-window algorithm for retrieving land
730 surface temperature from Landsat TM data and its application to the Israel-Egypt border region.
731 *International Journal of Remote Sensing, 22*, 3719-3746
732 Ramsey, M.S., & Flynn, I.T.W. (2020). The spatial and spectral resolution of ASTER infrared
733 image data: A paradigm shift in volcanological remote sensing. *Remote Sensing, 12*, 738
734 Sabol Jr, D.E., Gillespie, A.R., Abbott, E., & Yamada, G. (2009). Field validation of the
735 ASTER temperature–emissivity separation algorithm. *Remote sensing of environment, 113*,
736 2328-2344

737 Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P.,
738 Matricardi, M., & Geer, A. (2018). An update on the RTTOV fast radiative transfer model
739 (currently at version 12). *Geoscientific Model Development, 11*, 2717-2737
740 Silvestri, M., Romaniello, V., Hook, S., Musacchio, M., Teggi, S., & Buongiorno, M.F. (2020).
741 First comparisons of surface temperature estimations between ECOSTRESS, ASTER and
742 Landsat 8 over Italian volcanic and geothermal areas. *Remote Sensing, 12*, 184
743 Sobrino, J.A., Jiménez-Muñoz, J.C., Balick, L., Gillespie, A.R., Sabol, D.A., & Gustafson,
744 W.T. (2007). Accuracy of ASTER level-2 thermal-infrared standard products of an agricultural
745 area in Spain. *Remote sensing of environment, 106*, 146-153
746 Sobrino, J.A., & Skoković, D. (2016). Permanent stations for calibration/validation of thermal
747 sensors over Spain. *Data, 1*, 10
748 Townshend, J.R.G., Justice, C.O., Skole, D., Malingreau, J.P., Cihlar, J., Teillet, P., Sadowski,
749 F.a., & Ruttenberg, S. (1994). The 1 km resolution global data set: needs of the International
750 Geosphere Biosphere Programme. *International Journal of Remote Sensing, 15*, 3417-3441
751 Trigo, I.F., Ermida, S.L., Martins, J.P.A., Gouveia, C.M., Göttsche, F.-M., & Freitas, S.C.
752 (2021). Validation and consistency assessment of land surface temperature from geostationary
753 and polar orbit platforms: SEVIRI/MSG and AVHRR/Metop. *ISPRS Journal of*
754 *Photogrammetry and Remote Sensing, 175*, 282-297
755 Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface
756 temperature/emissivity product. *Remote sensing of environment, 140*, 36-45
757 Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface
758 temperature from space. *IEEE Transactions on Geoscience and Remote Sensing, 34*, 892-905
759 Wan, Z., & Li, Z.-L. (1997). A physics-based algorithm for retrieving land-surface emissivity
760 and temperature from EOS/MODIS data. *IEEE Transactions on Geoscience and Remote*
761 *Sensing, 35*, 980-996

762 Wang, K., & Liang, S. (2009). Evaluation of ASTER and MODIS land surface temperature
763 and emissivity products using long-term surface longwave radiation observations at
764 SURFRAD sites. *Remote sensing of environment, 113*, 1556-1565
765 Wang, M., He, C., Zhang, Z., Hu, T., Duan, S., Mallick, K., Li, H., Li, R., & Liu, X. (2022).
766 Evaluation of three land surface temperature products from Landsat series using in-situ
767 measurements. *under review*
768 Wang, M., Zhang, Z., Hu, T., & Liu, X. (2019). A practical single- channel algorithm for land
769 surface temperature retrieval: application to Landsat series data. *Journal of Geophysical*
770 *Research: Atmospheres, 124*, 299-316
771 Wang, M., Zhang, Z., Hu, T., Wang, G., He, G., Zhang, Z., Li, H., Wu, Z., & Liu, X. (2020).
772 An Efficient Framework for Producing Landsat-Based Land Surface Temperature Data Using
773 Google Earth Engine. *IEEE Journal of Selected Topics in Applied Earth Observations and*
774 *Remote Sensing, 13*, 4689-4701
775 Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen,
776 R.G., Anderson, M.C., Belward, A.S., & Cohen, W.B. (2019). Current status of Landsat
777 program, science, and applications. *Remote sensing of environment, 225*, 127-147
778 Yang, J., Zhou, J., Göttsche, F.-M., Long, Z., Ma, J., & Luo, R. (2020). Investigation and
779 validation of algorithms for estimating land surface temperature from Sentinel-3 SLSTR data.
780 *International Journal of Applied Earth Observation and Geoinformation, 91*, 102136
781 Yu, Y., Tarpley, D., Privette, J.L., Goldberg, M.D., Raja, M.K.R.V., Vinnikov, K.Y., & Xu, H.
782 (2008). Developing algorithm for operational GOES-R land surface temperature product. *IEEE*
783 *Transactions on Geoscience and Remote Sensing, 47*, 936-951
784 Yu, Y., Wang, H., Liu, Y., & Yu, P. (2017). Enterprise land surface emissivity algorithm
785 theoretical basis document. In: NOAA

786 Yuan, F., & Bauer, M.E. (2007). Comparison of impervious surface area and normalized
787 difference vegetation index as indicators of surface urban heat island effects in Landsat imagery.
788 *Remote sensing of environment, 106*, 375-386

789

Declaration of interests

☒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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: