

Generating daily 100 m resolution land surface temperature estimates continentally using an unbiased spatiotemporal fusion approach



Yi Yu ^{a,b,*}, Luigi J. Renzullo ^a, Tim R. McVicar ^c, Brendan P. Malone ^b, Siyuan Tian ^a

^a Fenner School of Environment & Society, The Australian National University, Canberra, ACT 2601, Australia

^b CSIRO Agriculture and Food, Canberra, ACT 2601, Australia

^c CSIRO Environment, Canberra, ACT 2601, Australia

ARTICLE INFO

Edited by Jing M. Chen

Keywords:

Land surface temperature
Spatiotemporal fusion
ESTARFM
Bias correction
MODIS
Landsat
ECOSTRESS

ABSTRACT

Fine spatial resolution (i.e., ≤ 100 m) land surface temperature (LST) data are crucial to study heterogeneous landscapes (e.g., agricultural and urban). Some well-known spatiotemporal fusion methods like the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) and the Enhanced STARFM (ESTARFM), which were originally developed to fuse surface reflectance data, may not be suitable for direct application in LST studies due to the high sub-diurnal dynamics of LST. Furthermore, the effectiveness of spatiotemporal fusion methods for LST data has not been thoroughly evaluated in previous studies that only focused on relatively small spatiotemporal extents. To address these limitations, we proposed a variant of ESTARFM, referred to as the unbiased ESTARFM (ubESTARFM), specifically designed to accommodate the high temporal dynamics of LST to generate fine-resolution LST estimates. We evaluated ubESTARFM and ESTARFM against in-situ LST and the ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) LST across 12 regions throughout Australia, encompassing various land covers and environments. Independent validation showed that ubESTARFM had a bias of 2.55 K, unbiased root mean squared error (ubRMSE) of 2.57 K, and Pearson correlation coefficient (R) of 0.95 against the in-situ LST over 11,290 observations at the 12 sites, all of which were considerably better than those calculated for ESTARFM, being a bias of 4.73 K, ubRMSE of 3.80 K and R of 0.92. When compared to ECOSTRESS data, ubESTARFM LST had a bias of -1.69 K, ubRMSE of 2.00 K, and R of 0.70 over 43 near clear-sky scenes, while ESTARFM LST had a bias of 1.79 K, ubRMSE of 2.68 K, and R of 0.59. Overall, our results demonstrated that ubESTARFM can avoid systematic bias accumulation, substantially reduce uncertainty deviation, and maintain a good level of correlation with validation datasets when compared to ESTARFM. A further assessment underscored the potential of ubESTARFM for application using LST data acquired from geostationary platforms (e.g., Himawari-8), with a mean ubRMSE (R) of 2.22 K (0.97) against in-situ LST over 1327 observations at 3 sites from southeast Australia at the overpass time of MODIS/Terra. This promising method leverages reliable numeric values from coarse-resolution LST while borrowing spatial heterogeneity from fine-resolution LST and has the potential to be coupled with energy balance and/or radiative transfer models thus enabling better farm and/or regional-scale water management strategies to be implemented. Furthermore, both the input and generated LST data, encompassing a comprehensive spatial extent over diverse land covers and climatic conditions, are publicly available for benchmarking future algorithmic refinements.

1. Introduction

Land surface temperature (LST) is an important geophysical variable, playing a key role in land-atmosphere exchanges and the surface radiation budget (Li et al., 2013). Knowledge of the spatiotemporal variations of LST provides information on the dynamics of the surface energy balance, which is fundamentally important in fields such as meteorology

and hydrology (Kerr et al., 2004; Li et al., 2013). As such, LST has a wide usage including drought monitoring (Hu et al., 2020; Zhang et al., 2017), retrieval of soil moisture and soil carbon (Abowarda et al., 2021; Long et al., 2019; Yu et al., 2021), estimation of evapotranspiration (ET) (Kalma et al., 2008; Long and Singh, 2013; Semmens et al., 2016), and serving as an indicator of climate change (Eleftheriou et al., 2018). Recently, there have been growing interests in fine spatial resolution (\leq

* Corresponding author at: Fenner School of Environment & Society, The Australian National University, Canberra, ACT 2601, Australia.

E-mail address: u6726739@anu.edu.au (Y. Yu).

100 m) LST due to its utility in agricultural applications (e.g., irrigation management and crop water stress assessment) and urban heat island studies, as the LST pixel size should be consistent with the spatial heterogeneity of individual fields in these studies (Anderson et al., 2008; Anderson et al., 2021; Ekinzog et al., 2022; Liu and Zhang, 2011; Yuan and Bauer, 2007).

However, the application potential of LST data is usually constrained due to a trade-off between the spatial and temporal resolutions of remotely sensed data, where a satellite sensor that can provide a fine spatial resolution usually has relatively poor temporal resolution, and vice versa (Emelyanova et al., 2013; Xia et al., 2019; Zhan et al., 2016). For example, the MODerate Resolution Imaging Spectroradiometer (MODIS) provides 1 km resolution LST data 4 times per day at maximum, whereas the Landsat 8 acquires 100 m resolution LST data but with a 16-day revisit time. Consequently, to overcome this dilemma, several fusion approaches to enhance both spatial and temporal resolutions of LST time series have been proposed (Chen et al., 2023; Gao et al., 2006; Wang and Atkinson, 2018; Zhu et al., 2010).

The spatiotemporal fusion approaches aim to predict fine-resolution data by integrating both time series and neighbourhood information from at least two satellite sensors. They produce synthesised data with both high spatial resolution and temporal frequency from low-frequency fine-resolution imagery (e.g., Landsat, Sentinel-2) and coarse-resolution high-frequency imagery (e.g., MODIS, Sentinel-3) (Zhu et al., 2016). One of the earliest and most extensively used approaches is the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) developed by Gao et al. (2006), which was originally implemented using surface reflectance from both Landsat and MODIS. The method assumes that temporal changes of reflectance within a coarse pixel are regarded as consistent. However, this assumption may be invalid across heterogeneous landscapes where coarse pixels have mixed land covers. Zhu et al. (2010) proposed a successor to STARFM, the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM), which uses a linear spectral mixing model to better account for landscape heterogeneity. ESTARFM shows an advancement across heterogeneous regions by considering geographic distance, spectral similarity and temporal correlation between coarse- and fine-resolution pixels (Emelyanova et al., 2013; Zhu et al., 2010). Both STARFM and ESTARFM can be described as weighting-based methods (Xia et al., 2019), which integrate the information from all inputs through weight functions. Other weighting-based fusion methods include the Flexible Spatiotemporal DAta Fusion (FSDAF) (Zhu et al., 2016), Spatial Temporal Adaptive Algorithm for mapping Reflectance CHange (STAARCH) (Hilker et al., 2009) and the Spatiotemporal Adaptive Data Fusion Algorithm for Temperature mapping (SADFAT) (Weng et al., 2014). The common feature of these fusion methods is the use of spatial relationship derived from fine- and coarse-resolution image pairs acquired at two times, to predict the fine-resolution image at a specific moment when only a coarse image is present.

There have been many studies that successfully applied these fusion methods to generate fine spatiotemporal resolution reflectance data (e.g., Emelyanova et al., 2013; Gevaert and García-Haro, 2015; Huang et al., 2013; Roy et al., 2008; Shen et al., 2016; Zhang et al., 2013). However, applications of these methods to generate fine-resolution LST data are relatively less explored (e.g., Long et al., 2020; Ma et al., 2022; Quan et al., 2018; Shi et al., 2022; Xia et al., 2019) and have three potential limitations. Firstly, the accuracy of LST retrieval algorithms is reliant on the precise determination of emissivity. This dependence on emissivity may lead to an exacerbation of landscape heterogeneity due to varying approaches for assigning emissivity values, resulting in a lack of comparability between LST data obtained from different sensors. Secondly, thermal properties are dynamic, so small differences (e.g., 60 min) in the overpass time of different satellites may lead to important differences in LST. Thirdly, and finally, the spatiotemporal fusion approaches cannot eliminate the systematic differences between sensors (Long et al., 2020). Hence, when fusing LST data from multiple sensors,

it is crucial to account for the comparability of the data, and to account for changes in the spatial and temporal patterns of LST distribution across the landscape that arise from variations in the rate of heating and cooling at different locations.

There are also some aspects in the experimental design of spatiotemporal fusion of LST data that can be improved. Table 1 summarises the details of recent studies that utilised STARFM, ESTARFM or FSDAF to generate LST data. A common limitation is that most studies were conducted with a limited spatial and temporal coverage, usually only a few thousand of km² over several months. Most LST-specific spatiotemporal fusion studies focused on built-up areas and bare soils (e.g., Li et al., 2021b; Liu et al., 2016; Shi et al., 2022; Xia et al., 2019) while few examined forested areas (e.g., Long et al., 2020; Zhu et al., 2021), hence most conclusions and implications may only be applicable to specific conditions. Nonetheless, Long et al. (2020) conducted the study within three 80 km × 80 km regions with some contrasting land covers and their objective was to downscale land surface modelled LST to the MODIS LST resolution (i.e., 1 km). Therefore, expanding the study regions and testing the algorithms over various landscapes and ecosystems could provide the opportunity for more comprehensive insights into the spatiotemporal LST fusion research. Furthermore, the usefulness of those spatiotemporal fusion approaches have not been fully evaluated previously as they were only compared to the original coarse-resolution data and in-situ data. Doing this does not provide fully independent verifications of the spatial details of the fused results, which, however, may have important impacts on further applications using LST data as input.

The objectives of this study are to: (1) modify the well-known ESTARFM algorithm to better accommodate spatiotemporal dynamics to generate LST data; (2) generate daily 100 m resolution LST data using both ESTARFM and the modified algorithm over 12 regions with various climates and landscapes at a continental scale of Australia between 2013 and 2021; (3) evaluate the generated LST against in-situ LST time series; and (4) validate the spatial patterns of generated LST with independent remotely sensed LST data.

2. Data and study area

The remotely sensed data and ground measurements used herein are itemised in Table 2. Landsat and MODIS LST data are used as inputs for the spatiotemporal fusion approach to generate fine-resolution LST, while ground longwave radiation and ECOSTRESS LST are respectively used for evaluation of the generated LST.

2.1. Remotely sensed data

2.1.1. MODIS LST

The MODIS sensor is onboard the Earth Observing System (EOS) Terra and Aqua satellites, launched in 1999 and 2002, respectively (Justice et al., 1998). They are in a sun-synchronous orbit with near-polar nominal descending node at ~10:30 local solar time (Terra) or ascending node at ~13:30 local solar time (Aqua). We collected time series of MOD11A1 data (Collection 6) over 01/Jan/2013–31/Dec/2021, which has a 1 km spatial resolution and daily temporal frequency, with overpass times at approximately 10:30 and 22:30 local solar time. LST retrievals are obtained using a generalised split-window algorithm with an accuracy around 2.0 K (Wan, 2014). MODIS data were downloaded from the National Aeronautics and Space Administration (NASA) Earthdata Search platform (<https://search.earthdata.nasa.gov/search>).

2.1.2. Landsat 8 LST

The Landsat 8 satellite, launched in 2013, carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments, and is in a sun-synchronous orbit. These instruments collect data along a 185-km swath with a 16-day revisit cycle and an equatorial crossing time of ~10:50 local solar time. The OLI measures in visible, near

Table 1

An overview of recent studies (i.e., since 2016) that used STARFM, ESTARFM, or FSDAF to generate LST. Studies are ordered alphabetically by the first author. Our study is added for completeness. The following abbreviations are used: ASTER: Advanced Spaceborne Thermal Emission and Reflection Radiometer; CLDAS: China Meteorological Administration Land Data Assimilation System; CLM 5.0: Community Land Model Version 5.0; HUTS: High-resolution Urban Thermal Sharpener.

Reference	Algorithm(s)	Study regions / validation area (km^2)	Landscape specifics	Input data	Study period	Key results for LST accuracy
Abowarda et al. (2021)	ESTARFM and HUTS	Haihe basin in north China / 9 *	Cropland, wetland, forestland and built-up areas	CLDAS, MODIS, Landsat 7 and Landsat 8	Mar 2015 - Oct 2017	30 m LST has R above 0.90, and bias, MAE, and RMSE all around 1 K.
Li et al. (2021b)	STARFM, ESTARFM and FSDAF	Zhangye, Gansu Province, China / ~ 1000	Farmland and Gobi	MODIS and Landsat 8	Apr 2013 - Nov 2016	ESTARFM yields better RMSE and R^2 (3.66 K and 0.92, respectively) than STARFM and FSDAF.
Liu et al. (2016)	STARFM	Beijing, China / ~ 7000	Built-up areas	MODIS and Landsat 8	May - Nov 2013	Landsat-like LST has RMSE above 1.36 K and R^2 above 0.71.
Liu and Weng (2018)	STARFM	Part of Los Angeles, California, USA / ~ 8000	Built-up areas, shrub, woodland and bare soils	MODIS and ASTER	Apr and Jul, 2007	LST differences were around 1 K from 15 to 1000 m resolution.
Long et al. (2020)	ESTARFM with final bias correction step	3 flux towers' surrounding areas in north China / 19,200	Irrigated cropland, forestland and sandy soils	CLDAS and MODIS	Early 2015 - Late 2017	MODIS-like LST estimates have MAE of 2.20–3.08 K, RMSE of 2.77–3.96 K, and R of 0.93–0.95.
Ma et al. (2022)	Filter-based spatiotemporal fusion model	Wuhan and Heihe River Basin, China / 1800	Impervious surfaces, cropland, bare land, grassland and water	CLM 5.0, MODIS and Landsat 7	Feb 2016 - Apr 2016	Sub-hourly Landsat-like LST has R of 0.96–0.99, MAE of 0.82–3.34 K, RMSE of 1.09–4.36 K.
Shi et al. (2022)	Modified FSDAF **	Beijing and Shenzhen, China / 3200	Built-up areas	MODIS and Landsat 5	2000–2003, 2014	Proposed method yields better results than STARFM and FSDAF in assessments using spatial accuracy indices.
Xia et al. (2019)	Weighted framework based on STARFM and a kernel-driven method	Beijing, China / ~ 1600	Built-up areas	MODIS and Landsat 8	Sep - Oct 2014, May - Sep 2017	Proposed framework has an improvement of 0.1–0.6 K than the kernel-driven method while has no remarkable difference with STARFM in accuracy.
Zhu et al. (2021)	Framework based on FSDAF and machine learning	Zhangye and Beijing, China / ~ 3200	Cultivated land, deserts, bare land, built-up areas and forestland	MODIS and Landsat 8	Several months in 2013, 2014 and 2017	Proposed framework yields better RMSE (0.85–2.29 K), relative RMSE (0.18–0.69 K) and R (all above 0.84) than three reference methods.
Our study	Modified ESTARFM with localised bias correction	12 flux towers' surrounding areas across Australia / 120,000	Forestland, cropland, shrubland, pasture and various soils	MODIS and Landsat 8	Jan 2013 - Dec 2021	Proposed method yields better bias, ubRMSE and R (2.55 K, 2.57 K and 0.95, respectively) than ESTARFM (4.73 K, 3.80 K and 0.92, respectively).

* The area of Haihe basin is 318,200 km^2 but they evaluated the spatial pattern of results using a 3 $\text{km} \times 3 \text{ km}$ subset.

** Modifications include considering mixed pixels, incorporating a new interpolation method, and using constrained least squares to combine spatiotemporal increments.

Table 2

Summary of data used herein. The spatial resolution of OzFlux measurements is not strictly a point, rather an aggregation of local area fluxes in the vicinity of the flux tower. This means the values of flux towers can be representative for a specific region, while how big the region is would depend on individual settings in the vicinity of each flux tower. The contribution from flux tower can range from meters to kilometres and can be considered comparable to remotely sensed data (Chu et al., 2021; Kljun et al., 2004).

Categories	Datasets	Variable	Spatial resolution	Temporal frequency	Period
Remotely sensed data	Landsat 8	LST	100 m	16-day	01/Jan/2013–31/Dec/2021
	MOD11A1	LST	1 km	Daily	01/Jan/2013–31/Dec/2021
	ECOSTRESS	LST	70 m	10-day	01/Jul/2018–31/Dec/2021
Ground measurements	OzFlux ground measurements	Longwave radiation	Point	30-min	01/Jan/2013–31/Dec/2021

infrared, and shortwave infrared bands with the spatial resolution of 30 m, while TIRS measures in two thermal infrared bands with the spatial resolution of 100 m. We collated the Landsat 8 level 1 data over 01/Jan/2013–31/Dec/2021. The Landsat 8 collection for Australia and its cloud masks are publicly available from the Digital Earth Australia (DEA; <https://dea.ga.gov.au/>). A split-window algorithm developed by Jiménez-Muñoz et al. (2014) was implemented to retrieve LST, with mean errors reported to be typically below 1.5 K. We utilised a Python package 'pylandtemp' developed by Mudele (2021) to conduct the LST retrieval.

However, previous studies reported the presence of stray light artifacts in Landsat 8 TIRS data, including banding patterns and absolute calibration discrepancies, which violate specified requirements in certain scenes (Montanaro et al., 2014; Reuter et al., 2015). These

artifacts are caused by out-of-field radiance that scatters onto the detectors, introducing a non-uniform signal across the field-of-view. Specifically, the effect is more pronounced in Band 11 compared to Band 10, with the magnitude typically being double in Band 11 (Barsi et al., 2014; Montanaro et al., 2014). To mitigate this issue, Montanaro et al. (2015) proposed an algorithm that employs near-coincident thermal data to establish per-detector functional relationships between incident out-of-field radiance and the additional stray light signal recorded by the TIRS detectors. However, further investigation is still required to assess the adequacy of this correction approach for operational usage (Cristóbal et al., 2018).

2.1.3. ECOSTRESS LST

The ECOSystem Spaceborne Thermal Radiometer Experiment on

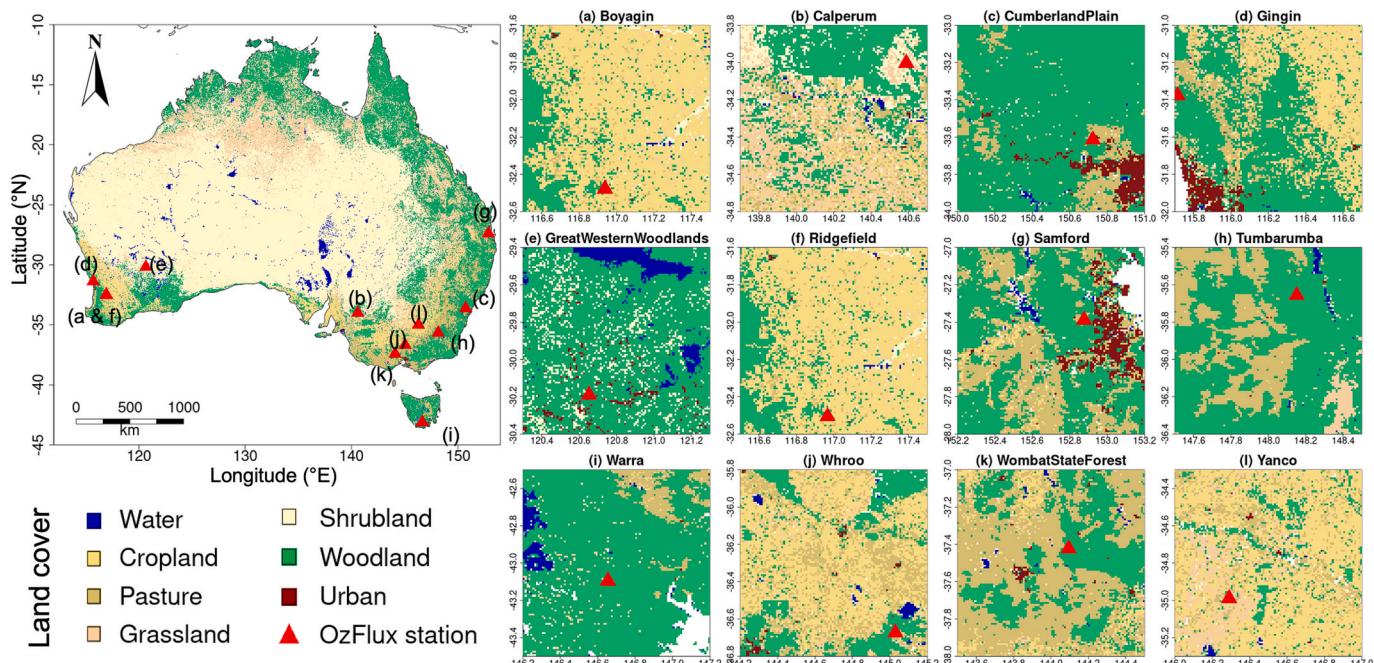


Fig. 1. The distribution of 12 OzFlux sites within Australia. (a-l) show the locations of flux towers within the ROIs and their land cover backgrounds (Lymburner et al., 2015). The size of each ROI is $1^\circ \times 1^\circ$. White areas in north-east of (g) Samford and south-east of (i) Warra are surrounding oceanic waters. Depending on the location of water features they can be dry salt-lake beds as common in central Australia.

Space Station (ECOSTRESS) mission is designed to measure plant temperature to explore how much water they need and how they respond to stress (Fisher et al., 2020). It was launched on 29/Jun/2018 by NASA and attached to the International Space Station (ISS). The satellite carries a TIR multispectral whiskbroom scanner with five bands between 8 and 12.5 μm (Fisher et al., 2020). The instrument scans in a ~ 400 km swath and allows considerable coverage of a targeted region with multiple revisits in a diurnal cycle. We collected the ECO2LSTE dataset (Version 1) over 01/Jul/2018–31/Dec/2021, which retrieves atmospherically corrected LST and emissivity using a physics-based Temperature/Emissivity Separation (TES) algorithm (Hook and Hulley, 2019). The spatial resolution of these data is 70 m and needs to be co-processed with corresponding ECO1BGEO and ECO2CLD datasets that provide relevant geolocation information and cloud mask, respectively. The ECOSTRESS LST has a RMSE of ~ 1 K against ground-based measurements (Hulley et al., 2021). We acquired the ECO2LSTE, ECO1BGEO and ECO2CLD datasets from the NASA Earthdata Search platform (<https://search.earthdata.nasa.gov/search>).

2.2. Study area and ground measurements

This study is conducted on twelve $1^\circ \times 1^\circ$ tiles (i.e., each of the 12 tiles is $\sim 10,000 \text{ km}^2$) located across the Australian continent. Within each tile a well-maintained flux tower site is located that forms part of the OzFlux network (Beringer et al., 2016). Fig. 1 shows the distribution of the flux sites across Australia and their land cover types (Lymburner et al., 2015) within a region of interest (ROI). Details of the 12 chosen flux sites are provided in Table 3. We divided the sites into 4 climate classification groups, including semi-arid, humid subtropical, oceanic and cool temperate. Furthermore, based on the land cover information, we also classified the Boyagin, Ridgefield and Yanco site as non-vegetated, and the other 9 sites as vegetated.

We used the observed upwelling and downwelling longwave radiation (level 3) measured at each of the 12 sites, and by rearranging the standard longwave radiation balance equation (Allen et al., 1998) were able to estimate a 'ground-based' LST as:

$$T_s = \left(\frac{F^\downarrow - (1 - \epsilon_b) F^\downarrow}{\sigma \epsilon_b} \right)^{1/4} \quad (1)$$

Table 3

Summary information for each of the selected flux tower sites. The climate classification is based on a Köppen-Geiger climate classification (Beck et al., 2018), and information for both 'Climate classification' and 'Land cover' columns are obtained from <https://ozflux.org.au/>.

Site name	Latitude ($^\circ\text{N}$)	Longitude ($^\circ\text{E}$)	Landsat 8 path / row tile(s)	Climate classification	Land cover
Boyagin	-32.477	116.939	112/082	Semi-arid	Woodland and cropland
Calperum	-34.003	140.588	096/084	Semi-arid	Recovering woodland
Cumberland Plain	-33.615	150.724	090/083	Humid subtropical	Dry sclerophyll
Gingin	-31.376	115.714	112/082	Oceanic	Coastal health woodland
Great Western Woodlands	-30.191	120.654	110/080, 110/081	Semi-arid	Temperate woodland, shrubland and mallee
Ridgefield	-32.506	116.967	112/082	Semi-arid	Dryland agriculture
Samford	-27.388	152.877	089/079	Humid subtropical	Improved pasture
Tumbarumba	-35.657	148.152	091/085	Oceanic	Wet temperate sclerophyll eucalypt
Warra	-43.095	146.655	090/090	Cool temperate	Eucalyptus forest
Whroo	-36.673	145.029	093/085	Cool temperate	Box woodland
Wombat State Forest	-37.422	144.094	093/086	Cool temperate	Dry sclerophyll eucalypt forest
Yanco	-34.989	146.291	092/084	Semi-arid	Various soils and cropland

where T_s denotes LST (K), F^\dagger is the upwelling longwave radiation (W/m^2), F^\ddagger is the downwelling longwave radiation (W/m^2), e_b is the surface broadband emissivity (unitless), and σ represents the Boltzmann constant ($5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$). We accessed the broadband emissivity from the University of Wisconsin (UW) Baseline Fit Emissivity Database, a global database of land surface emissivity derived using MODIS operational land surface emissivity (Seemann et al., 2008). We calculated the monthly emissivity by taking the average of the values from 10.8- and 12.1- μm bands for each month, then used the monthly value of emissivity, along with the longwave radiation, to derive the ground-based LST every 30-min.

3. Methodology

3.1. ESTARFM

The ESTARFM algorithm assumes that remotely sensed data from different sensors are acquired at a close time and so are comparable and correlated with each other after undergoing radiometric calibration and appropriate pre-processing (Zhu et al., 2010). It has been extensively used and demonstrated robustness in surface reflectance studies (Emenyanova et al., 2013), with only a few studies reporting its usage in LST research (Long et al., 2020; Yang et al., 2016).

The ESTARFM algorithm begins with a search for the pixels with similar values within a moving window, where similarity is defined as:

$$|F(x_i, y_i, t_0) - F(x_{w/2}, y_{w/2}, t_0)| \leq 2 \times \frac{\sigma(P)}{n} \quad (2)$$

where $F(x_i, y_i, t_0)$ is the i th pixel within the searching window of $w \times w$ pixels at time t_0 from the fine-resolution imagery; $F(x_{w/2}, y_{w/2}, t_0)$ is the central pixel $(x_{w/2}, y_{w/2})$ of the fine-resolution imagery within the same geographic area of the search window at time t_0 ; $\sigma(P)$ is the standard deviation of a given patch P ; and n is the number of land cover classes.

If a pixel satisfies Eq. (2), it will be classified as a spectrally similar homogeneous pixel with the central pixel within the window then be used to calculate weights and coefficients. The central pixel of the moving window at another time t_p is predicted as:

$$F(x_{w/2}, y_{w/2}, t_p) = F(x_{w/2}, y_{w/2}, t_0) + \sum_{i=1}^N W_i \times V_i \times (C(x_i, y_i, t_p) - C(x_i, y_i, t_0)) \quad (3)$$

where N is the number of similar pixels within the search window; W_i is the weight of the i th similar pixel; V_i is a regression coefficient between the chosen fine- and coarse-resolution pixels; $C(x_i, y_i)$ is the pixel at the location (x_i, y_i) of a coarse-resolution image. W_i is given as:

$$W_i = (1/D_i) / \sum_{i=1}^N (1/D_i) \quad (4)$$

$$D_i = (1 - R_i) \times d_i \quad (5)$$

$$d_i = 1 + \sqrt{(x_{w/2} - x_i)^2 + (y_{w/2} - y_i)^2} / (w/2) \quad (6)$$

where D_i is an index combining the spectral and spatial similarity; d_i is the distance between the i th pixel and the central pixel within the window; R_i is the spectral correlation coefficient between fine- and coarse-resolution pixels for the i th similar pixel.

The final prediction is obtained by a weighted combination of the two individual predictions:

$$F(x_{w/2}, y_{w/2}, t_p) = T_m \times F_m(x_{w/2}, y_{w/2}, t_p) + T_n \times F_n(x_{w/2}, y_{w/2}, t_p) \quad (7)$$

$$T_k = \frac{1 / \left| \sum_{j=1}^w \sum_{l=1}^w C(x_j, y_l, t_k) - \sum_{j=1}^w \sum_{l=1}^w C(x_j, y_l, t_p) \right|}{\sum_{k=m,n} \left(1 / \left| \sum_{j=1}^w \sum_{l=1}^w C(x_j, y_l, t_k) - \sum_{j=1}^w \sum_{l=1}^w C(x_j, y_l, t_p) \right| \right)}, \quad (k \\ = m, n) \quad (8)$$

where T_k ($k \in (m, n)$) are the temporal weights of the individual predictions.

3.2. Unbiased ESTARFM (ubESTARFM)

Initially developed for surface reflectance data, ESTARFM assumes that fine- and coarse-resolution data are comparable and aims to minimize systematic biases. However, in LST studies, fine- and coarse-resolution data might not always be comparable due to differences in their retrieval algorithms, potential anomalies in multi-sourced emissivity and differences in observation angles. LST is sensitive to temporal variations, and even a small difference in the acquisition time (e.g., 60 min) may lead to substantial and non-linear difference in estimated LST values, while the overpass time of Landsat and MODIS are usually not considered equivalent. Moreover, the prediction accuracy of ESTARFM is highly dependent on the exact value of fine-resolution data, along with the temporal variation of coarse-resolution data (Eq. (3)), hence it is crucial to use highly accurate fine-resolution LST data. Time series-based bias correction is an extensively employed approach in hydrology and climate studies (Chen et al., 2011; Leander and Buishand, 2007; Long et al., 2020). However, the direct application of this technique in Landsat-MODIS LST fusion may not be applicable due to the sparse temporal frequency of Landsat LST data (i.e., only 1/16 of the observation frequency of MODIS), which renders the characteristics of Landsat LST time-series less representative for generating daily results. Furthermore, the considerable absolute calibration discrepancies in Landsat-8 TIRS data (Barsi et al., 2014; Montanaro et al., 2014; Reuter et al., 2015) precludes the use of fine-resolution time series data as a reference. Therefore, by integrating this approach and localised spatial information, we propose a variant of ESTARFM here, referred to as the unbiased ESTARFM (ubESTARFM), to better accommodate spatiotemporal fusion to LST data.

In ubESTARFM, we applied a local bias correction on the central pixel and similarly fine-resolution pixels within the window using the mean value of corresponding coarse-resolution pixels as a reference:

$$F(x_{w/2}, y_{w/2}, t_0)^{BC} = F(x_{w/2}, y_{w/2}, t_0) - \mu(F(x_i, y_i, t_0)) + \mu(C(x_i, y_i, t_0)) \quad (9)$$

$$F(x_i, y_i, t_0)^{BC} = F(x_i, y_i, t_0) - \mu(F(x_i, y_i, t_0)) + \mu(C(x_i, y_i, t_0)) \quad (10)$$

where $F(x_{w/2}, y_{w/2}, t_0)^{BC}$ and $F(x_i, y_i, t_0)^{BC}$ are the central pixel within the window and similarly fine-resolution pixels at t_0 after bias correction, respectively; $\mu(F(x_i, y_i, t_0))$ is the mean of similarly fine-resolution pixels and $\mu(C(x_i, y_i, t_0))$ is the mean of the corresponding coarse-resolution pixels. By applying this linear scaling approach, we can scale the systematic biases of the fine-resolution data to a same level of the corresponding coarse-resolution data in each moving window, while keeping the variation and spatial details of fine-resolution data. Fig. 2 illustrates how ubESTARFM works. Hence, the R_i and W_i are accordingly modified and we can incorporate the bias corrected pixels into the prediction then modified Eq. (3) as:

$$F(x_{w/2}, y_{w/2}, t_p) = F(x_{w/2}, y_{w/2}, t_0)^{BC} + \sum_{i=1}^N W_i \times V_i \times (C(x_i, y_i, t_p) - C(x_i, y_i, t_0)) \quad (11)$$

This localised bias correction strategy can work optimally when coarse-resolution LST demonstrates better performance than fine-

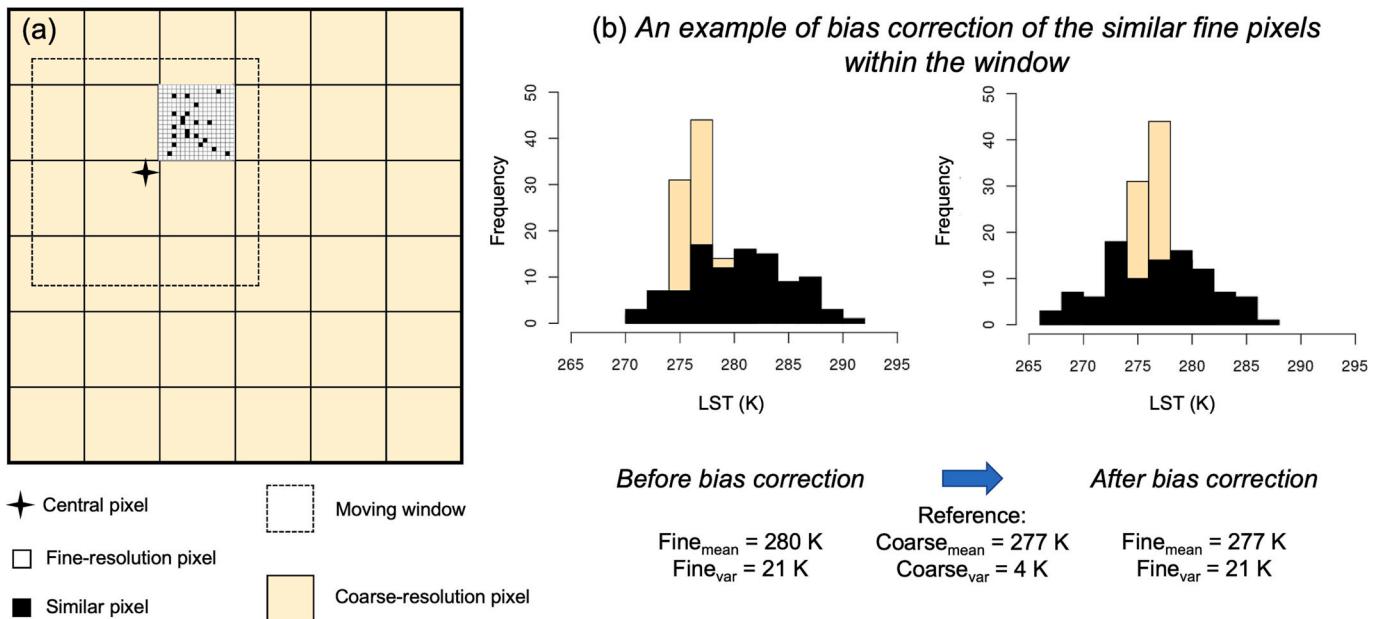


Fig. 2. A schematic to show how ubESTARFM works. Part (a) gives an example of the distribution of the similar fine-resolution pixels and corresponding coarse-resolution pixels within a searching window of ubESTARFM, and part (b) shows the distribution of them after bias correction using coarse-resolution pixels as reference. In both plots that comprise part (b) the black columns represent the similar fine-resolution pixels and the light-yellow columns represent corresponding coarse-resolution pixels.

resolution LST. Moreover it is also expected to be effective under the condition that coarse-resolution LST data exhibits comparable or even slightly degraded agreement with reference LST data (typically obtained from in-situ measurements) when compared to fine-resolution LST data for three main reasons. Firstly, the prediction strategy of ESTARFM (Eq. (3)) relies on an integration of fine-resolution data values and the difference in coarse-resolution data between the prediction and training dates. Assuming fine-resolution data demonstrates slightly superior performance compared to coarse-resolution data, there is no prior knowledge regarding whether both fine- and coarse-resolution data exhibit the same positive or negative bias performance. A direct integration of multi-sourced data without addressing this issue is likely to accumulate errors. Secondly, the coarse-resolution LST value on a prediction date, denoted as $C(x_i, y_i, t_p)$, plays a crucial role in generating daily outputs and is typically irreplaceable, as corresponding fine-resolution data values are unavailable on prediction dates for bias correction. Therefore, it is reasonable to choose the coarse-resolution data as the reference for bias-correcting the fine-resolution LST on a training date (Eq. (9)) to ensure consistency in uncertainty levels. Thirdly, even in cases where fine-resolution data is superior and could be utilised as the reference for localised bias correction, it would yield equivalent results to our current strategy:

$$F(x_{w/2}, y_{w/2}, t_p) = F(x_{w/2}, y_{w/2}, t_0) + \sum_{i=1}^N W_i \times V_i \times (C(x_i, y_i, t_p) - C(x_i, y_i, t_0)^{BC}) \quad (12)$$

$$C(x_i, y_i, t_0)^{BC} = C(x_i, y_i, t_0) - \mu(C(x_i, y_i, t_0)) + \mu(F(x_i, y_i, t_0)) \quad (13)$$

By moving the term $C(x_i, y_i, t_0)^{BC}$ out of the bracket, the predicted result would still undergo a weighted removal of $\mu(F(x_i, y_i, t_0))$ and a weighted addition of $\mu(C(x_i, y_i, t_0))$. However, it is anticipated that this strategy, referred to as ubESTARFM_{fine}, may be sub-optimal as it retains residuals of systematic biases from both data sources.

3.3. Evaluation methods

To minimize the influence of sub-diurnal LST dynamics, we need to ensure the acquisition times of remotely sensed data are comparable with ground measurements. The overpass times of MODIS satellites are not constant, ranging from 10:00 to 12:10 local solar time for Terra (Hu et al., 2014). To make the acquisition times of MODIS LST consistent with the records of in-situ LST, we convert the 'view time' layer (which is in local solar time) in MODIS LST data to the GMT format, which is given as:

$$T_{view,GMT} = T_{view,solar} - Lon / 15 \quad (14)$$

where $T_{view,GMT}$ is the MODIS view time in GMT time zone, $T_{view,solar}$ is the MODIS view time in local solar time, and Lon is the longitude in decimal of a given pixel.

Then we use a one-hour window to search for corresponding in-situ LST, as OzFlux sites measure longwave radiation every 30 min:

$$T_{view,GMT} - 30\text{ min} \leq T_{in situ,GMT} \leq T_{view,GMT} + 30\text{ min} \quad (15)$$

where $T_{in situ,GMT}$ is the time of in-situ LST in GMT time zone. We calculate a mean value if there are two in-situ records within this temporal window. Implementing this means we compare MODIS LST with ground measurements within a ± 30 -min window regardless of their individual time zones. Furthermore, we evaluated the applicability of this strategy by comparing the in-situ LST values from two temporal windows (i.e., 10:00–10:30 and 10:30–11:00 local standard time, respectively) against their mean values. The calculated bias and unbiased root mean square error (ubRMSE) among 30,900 samples are ± 0.64 K and 0.52 K, respectively. The detailed statistics can be found in Table S1. These findings unequivocally indicate a marginal deviation and affirm the feasibility of adopting this strategy.

3.3.1. Evaluation metrics

We calculated the following metrics to evaluate the performance of fused results against in-situ data, including bias and ubRMSE:

$$Bias = \mu(LST_{fuse}) - \mu(LST_{ref}) \quad (16)$$

$$ubRMSE = \sqrt{\frac{\sum ((LST_{fuse} - \mu(LST_{fuse})) - (LST_{ref} - \mu(LST_{ref})))^2}{N}} \quad (17)$$

where LST_{fuse} is the time series of fused LST; LST_{ref} is the time series of reference LST (i.e., in-situ LST); N is the number of individual observations in the time series.

3.3.2. Cross-satellite comparison

The Landsat and MODIS/Terra mission have sun-synchronous orbits and relatively stable overpass times. However, the overpass time of ECOSTRESS varies daily due to the asynchronous orbit of the ISS. Hence, we used the same strategy as proposed in Eq. (13) to search for a temporally close ECOSTRESS scene with MODIS LST, i.e., we implemented a ± 30 -min window to match the overpass time of ECOSTRESS and MODIS LST. Furthermore, we apply the metrics in Eqs. (14) and (15), and a Pearson correlation coefficient on a pixel-wise basis:

$$R = \frac{\sum (LST_{fuse} - \mu(LST_{fuse})) (LST_{ref} - \mu(LST_{ref}))}{\sqrt{\sum (LST_{fuse} - \mu(LST_{fuse}))^2 \sum (LST_{ref} - \mu(LST_{ref}))^2}} \quad (18)$$

where R is the Pearson correlation coefficient; LST_{ref} is the reference LST (i.e., ECOSTRESS LST).

3.4. Experimental design

Fig. 3 shows the schematic of this study. Firstly, we performed a temporal evaluation of the Landsat LST on every 16 days and the corresponding MODIS LST using the in-situ LST data, to evaluate if the

coarse-resolution LST at least has comparable agreement with in-situ data compared to the fine-resolution LST. Secondly, we paired Landsat LST and MODIS LST every 16 days (i.e., on the dates when Landsat data were acquired) for a given ROI between 1/Jan/2013 and 31/Dec/2021. The completeness of training scenes plays a crucial role in obtaining accurate results using (ub)ESTARFM, as one missing pixel in training pairs can lead to the absence of that pixel throughout the entire time series and affect the prediction accuracy within a moving window (see Eq. (3)). Thus, we shortlisted the pairs by only considering those Landsat and MODIS scenes with $>2/3$ clear-sky pixels after applying their respective cloud masks. After that, we trained both ESTARFM and ubESTARFM on every two pairs of the shortlisted LST candidates, then performed the prediction using the MODIS LST acquired between the two pairs. The patch size is 200×200 pixels and the moving window is 51×51 pixels, which follow the default settings of (Zhu et al., 2010). Ma et al. (2018) demonstrated that augmenting the size of the moving window generally leads to enhanced accuracy; while this enhancement diminishes as the window size exceeds a certain threshold, potentially resulting in a decline in accuracy. Through a series of iterative experiments, we found that the accuracy of the results was insensitive to variations in window size beyond a threshold of approximately two thousand pixels, hence we retained the default settings of those parameters. Using both ESTARFM and ubESTARFM, we obtained two datasets of fused LST, both of which have a 100 m spatial resolution and a daily frequency (cloud permitting). Thirdly and finally, we performed a temporal evaluation of the fused LST against the in-situ LST and a spatial comparison with the ECOSTRESS LST.

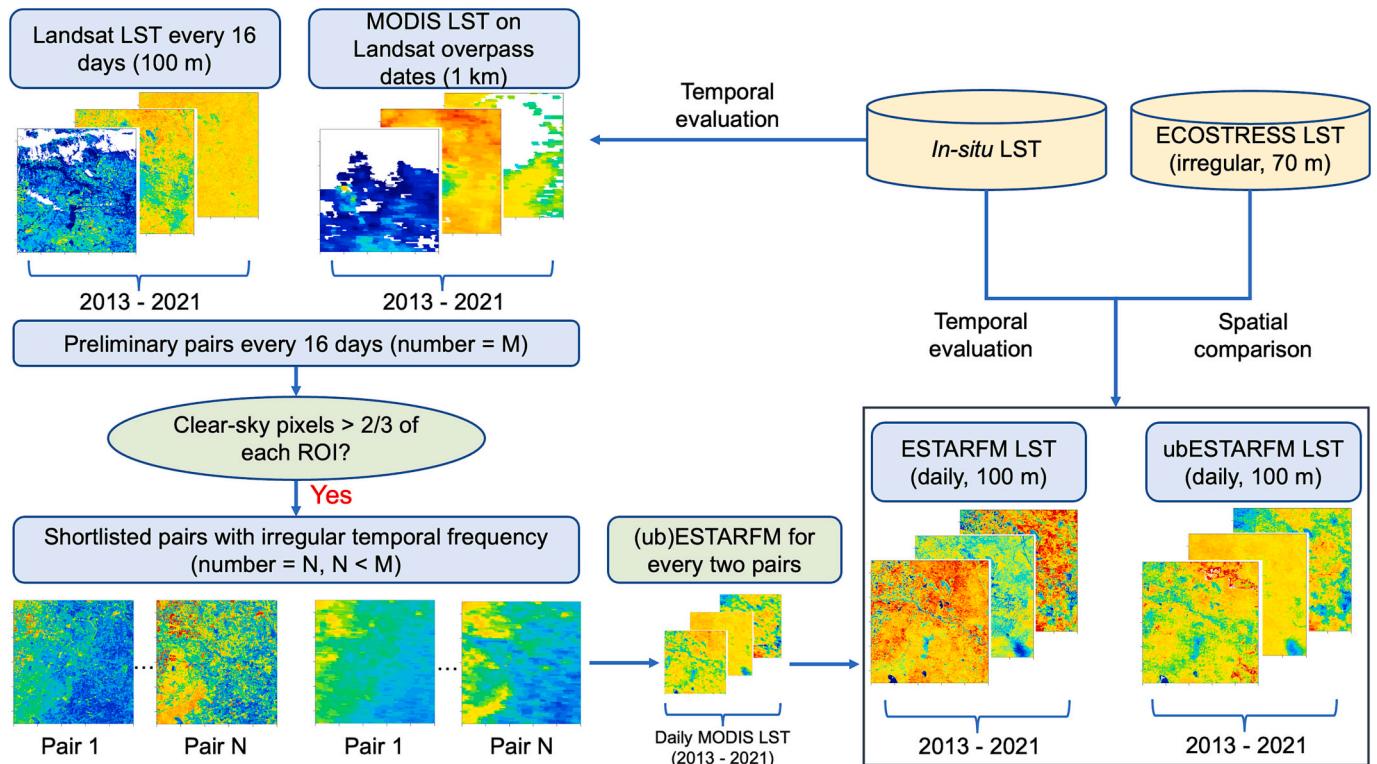


Fig. 3. The framework diagram of this study. The Landsat LST and MODIS LST acquired on the same date (i.e., Landsat overpass date) are deemed 'a pair'. M denotes the total number of Landsat-MODIS pairs with a temporal frequency of 16-days spanning the period from 01/Jan/2013 to 31/Dec/2021; N represents the number of pairs that have been shortlisted, with an irregular frequency, based on a clear-sky ratio exceeding $2/3$ for both images. Both ESTARFM and ubESTARFM were implemented on every two consecutive image-pairs. For example, the first iteration employed the first and second Landsat-MODIS pairs for training and subsequently made predictions using the daily MODIS LST within this specific period. The second iteration used the second and third Landsat-MODIS pairs, and so forth. This framework was independently implemented at each ROI.

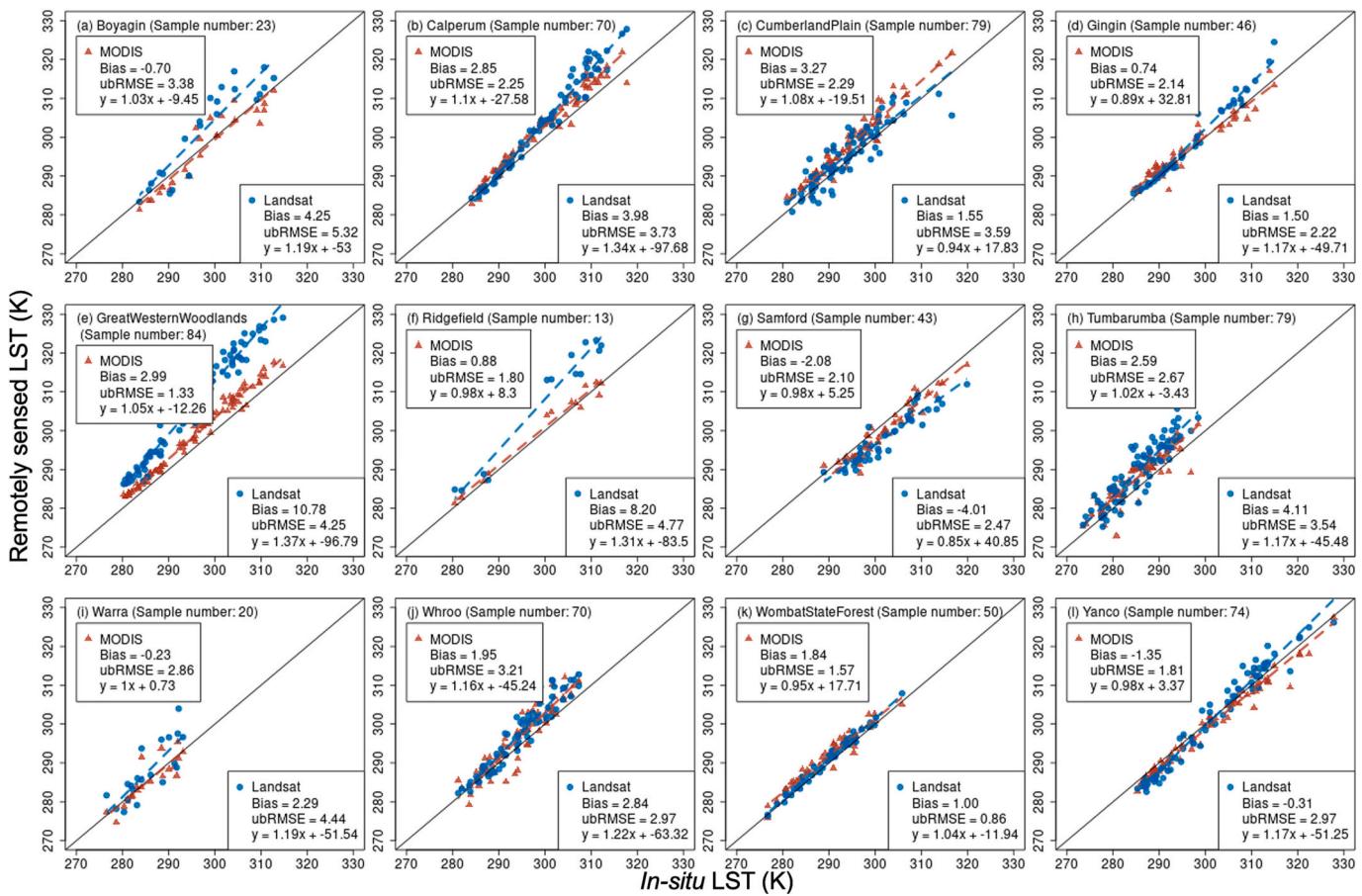


Fig. 4. Scatterplots of Landsat LST and MODIS LST against in-situ LST across the 12 sites when both data are available. The units of the Bias and ubRMSE statistics shown on each plot are K. The thin black continuous diagonal line on each plot is the 1:1 line. The red and blue dashed lines are the linear regression lines of best-fit (as provided by the equations on each sub-part) for MODIS and Landsat, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Results

4.1. Evaluation of Landsat and MODIS LST against in-situ LST

Fig. 4 shows the scatterplots of Landsat LST and corresponding MODIS LST against in-situ LST at all 12 sites over the study period (i.e., 1/Jan/2013 to 31/Dec/2021). We performed the assessment at the native resolution of MODIS and Landsat (i.e., 1 km and 100 m, respectively). This is because we found Landsat LST showed similar performance against in-situ LST at 100 m and after aggregated to 1 km resolution (Table S2). There are at least 20 cloud-free samples at all sites over these 9 years except for the Ridgefield site. In terms of bias, MODIS LST outperforms Landsat LST at nine sites, except for the Cumberland Plain, Wombat State Forest, and Yanco sites. In terms of ubRMSE, MODIS LST outperforms Landsat LST at 10 sites, except for the Whroo and Wombat State Forest sites. The best performance of bias for MODIS LST (-0.23 K) is observed at the Warra site (Fig. 4i), which is covered by forests, while the best ubRMSE (1.33 K) is observed at the Great Western Woodlands site (Fig. 4e). The best performance of bias for Landsat LST (-0.31 K) is observed at the Yanco site (Fig. 4l) with a land cover of cropland, while the best ubRMSE (0.86 K) is observed at the Wombat State Forest site (Fig. 4k). In general, Landsat LST exhibits more uncertainty, such as the observed substantial bias at the Great Western Woodlands site despite having a relatively moderate deviation (Fig. 4e), though it demonstrates superior agreement with in-situ LST at the Wombat State Forest site (Fig. 4k). By comparison, MODIS LST exhibits stable performance at all sites.

Table 4 summarises the metrics comparing MODIS and Landsat LST

Table 4

Evaluation statistics of Landsat LST and MODIS LST across the 12 sites and as grouped into climate classifications (semi-arid, humid subtropical, oceanic and cool temperate) and land cover classes (vegetated and non-vegetated) during 1/Jan/2013–31/Dec/2021. The units of the Bias and ubRMSE statistics are K.

Group name	Bias		ubRMSE		Sample number
	MODIS	Landsat	MODIS	Landsat	
Semi-arid	1.31	5.17	1.91	3.87	264
Humid subtropical	1.38	-0.41	2.22	3.19	122
Oceanic	1.91	3.15	2.47	3.06	125
Cool temperate	1.60	2.10	2.57	2.43	140
Vegetated	2.00	3.37	2.24	3.20	541
Non-vegetated	-0.95	1.65	2.14	3.68	110
All	1.50	3.08	2.22	3.28	651
11 sites *	1.51	2.97	2.23	3.25	638

* Excluding the Ridgefield site, which has too few samples for significant statistics.

against in-situ LST at 12 sites. Across different climate classifications, the bias of MODIS LST is relatively stable, ranging from 1.31 K to 1.91 K, whereas the bias of Landsat LST ranges from -0.41 K to 5.17 K; the ubRMSE of MODIS LST ranges from 1.91 K to 2.57 K, while that of Landsat LST ranges from 2.43 K to 3.87 K. Under varying vegetation conditions, both MODIS and Landsat LST have better bias at non-vegetated sites (-0.95 K and 1.65 K, respectively) than the vegetated sites (2.00 K and 3.37 K, respectively); the ubRMSE of MODIS LST is 2.24 K at vegetated sites and 2.14 K at non-vegetated sites, while that of

Landsat LST is 3.20 K and 3.68 K, respectively. The overall bias and ubRMSE of MODIS LST against in-situ LST are 1.50 K and 2.22 K, respectively. In comparison, the overall bias and ubRMSE of Landsat LST against in-situ LST are 3.08 K and 3.28 K, respectively. When excluding the Ridgefield site, the performance gap between MODIS and Landsat LST is slightly narrowed, with a bias of 1.51 K and 2.97 K, respectively, and a ubRMSE of 2.23 K and 3.25 K, respectively. Overall, the results indicate that MODIS LST has a better agreement with in-situ LST than Landsat LST under most conditions, indicating the suitability of coarse-resolution LST to be used as a reference in the spatiotemporal fusion process.

4.2. Assessment of algorithm performance

We compared the fitting performance of two algorithms using MODIS LST as reference. Fig. 5 shows scatterplots comparing the results of ESTARFM and ubESTARFM against the MODIS LST at the 12 sites during the training dates. The ubESTARFM LST exhibits superior agreement with the MODIS LST at all sites, with substantially less uncertainty when compared to the ESTARFM LST. The ubESTARFM LST performs better in terms of bias than the ESTARFM LST at 11 sites, except for Whroo. Additionally, the ubESTARFM LST outperforms the ESTARFM LST in ubRMSE at all 12 sites. Notably, at sites where ESTARFM predictions exhibit high bias, such as the Boyagin (Fig. 5a), Great Western Woodlands (Fig. 5e), and Ridgefield (Fig. 5f), the ubESTARFM substantially improves agreement between the prediction and MODIS LST.

Table 5 presents a summary of the metrics for the results from

ESTARFM and ubESTARFM against the MODIS LST at the training dates. The ubESTARFM outperforms ESTARFM in terms of ubRMSE, with a range of 0.64 K to 1.27 K compared to ESTARFM's range of 2.20 K to 3.92 K across all conditions. When it comes to bias performance under different climates, ubESTARFM performs remarkably better under semi-arid and oceanic climates (both 0.13 K) than ESTARFM (2.56 K and 3.33 K, respectively). Though ESTARFM has an almost perfect bias performance under humid subtropical climate (-0.01 K) whereas that of ubESTARFM is -1.20 K, they are comparable under cool temperate climate (-0.84 K and -0.87 K for ESTARFM and ubESTARFM, respectively). Under varying vegetation conditions, they have comparable bias at vegetated sites (0.48 K and -0.52 K for ESTARFM and ubESTARFM, respectively). However, ubESTARFM has a remarkably better performance of bias at non-vegetated sites (0.20 K) than ESTARFM (4.29 K). The overall bias and ubRMSE for ubESTARFM LST are -0.34 K and 0.90 K, respectively, whereas those for ESTARFM LST are 1.43 K and 3.27 K, respectively. If the Warra site is excluded, the metrics for ubESTARFM LST change slightly to -0.35 K and 0.88 K, respectively, whereas those for ESTARFM LST change to 1.40 K and 3.24 K, respectively. In general, ubESTARFM reduces the uncertainty level in prediction by about 75% when compared to ESTARFM output.

We showcase the difference in spatial patterns of the algorithms on the training dates at the Calperum and Samford sites (Fig. 6). At Calperum (Fig. 6a-h), the MODIS and Landsat LST on 10/Nov/2014 have a difference of approximately 5 K. The ESTARFM LST (Fig. 6c) is identical to the Landsat LST (Fig. 6b), while the ubESTARFM LST (Fig. 6d) has a similar pattern and value scale to the MODIS LST (Fig. 6a), but with sharper details. In the 10 km × 10 km zoom window, it is observed that

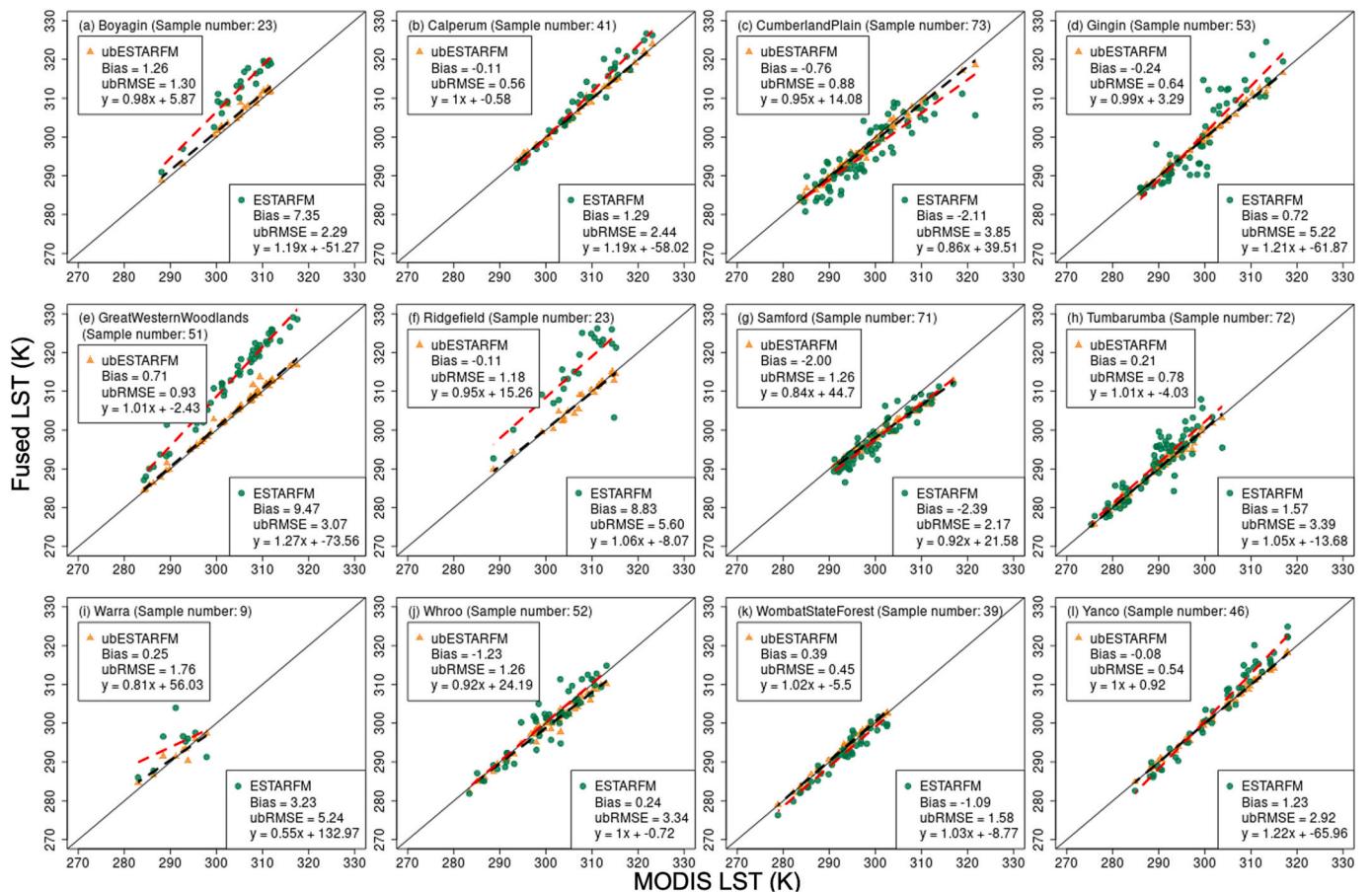


Fig. 5. Scatterplots of ESTARFM and ubESTARFM LST against MODIS LST at the 12 sites on the training dates. The units of the Bias and ubRMSE statistics are K. The thin black continuous diagonal line on each plot is the 1:1 line. The red and black dashed lines are the linear regression lines of best-fit (as provided by the equations on each sub-part) for ESTARFM and ubESTARFM, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5

Statistics of ESTARFM and ubESTARFM LST against MODIS LST across the 12 sites and as grouped into climate classifications (semi-arid, humid subtropical, oceanic and cool temperate) and land cover classes (vegetated and non-vegetated) on the training dates. The units of the Bias and ubRMSE statistics are K.

Group name	Bias		ubRMSE		Sample number
	ESTARFM	ubESTARFM	ESTARFM	ubESTARFM	
Semi-arid	2.56	0.13	3.17	0.64	230
Humid subtropical	-0.01	-1.20	2.20	1.27	94
Oceanic	3.33	0.13	3.92	0.88	95
Cool temperate	-0.84	-0.87	3.74	1.09	134
Vegetated	0.48	-0.52	3.42	0.96	415
Non-vegetated	4.29	0.20	2.83	0.69	138
All	1.43	-0.34	3.27	0.90	553
11 sites *	1.40	-0.35	3.24	0.88	544

* Excluding the Warra site, which has too few samples for significant statistics.

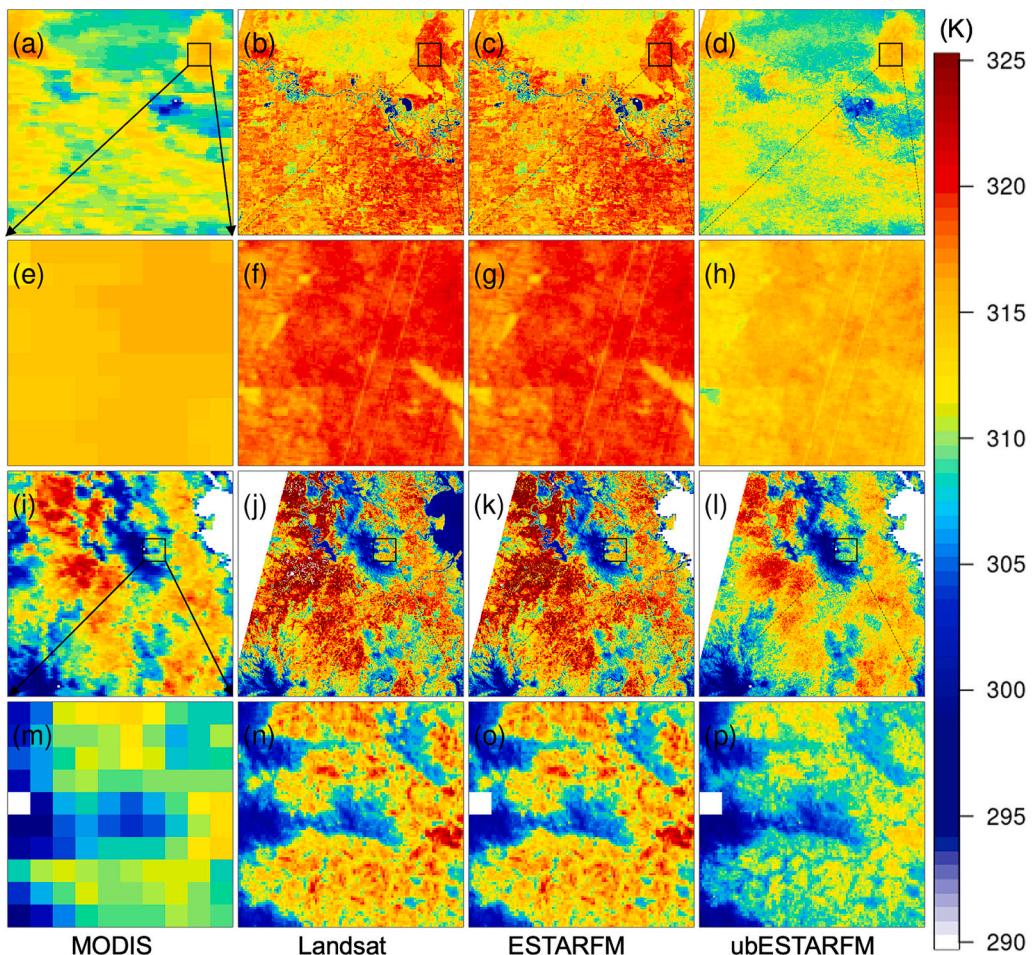


Fig. 6. Two examples from the Calperum and Samford site to show the difference between MODIS, Landsat, ESTARFM and ubESTARFM LST on the training dates. (a-d) are the LSTs within a homogenous $1^\circ \times 1^\circ$ ROI (Calperum) on 10/Nov/2014, and the black window is a $10\text{ km} \times 10\text{ km}$ area surrounding the in-situ site; (e-h) are zoomed areas of the window in (a-d); (i-l) are the LSTs within a heterogeneous ROI (Samford) on 29/Nov/2016, with the black window representing the surrounding area of the in-situ site; (m-p) are zoomed areas of the window in (i-l).

the ESTARFM LST (Fig. 6g) is still identical to the Landsat LST (Fig. 6f); however, the ubESTARFM LST (Fig. 6h) has a similar value scale with the MODIS LST (Fig. 6e), and also provides similar spatial details to the Landsat LST over this homogenous region. At the relatively heterogeneous Samford site (Fig. 6i-p) on 29/Nov/2016, the ESTARFM LST (Fig. 6k, o) is still identical to the Landsat LST (Fig. 6j, n) at all scales. In comparison, the ubESTARFM LST (Fig. 6l) has a good consistency with the MODIS LST (Fig. 6i) within the $1^\circ \times 1^\circ$ ROI. In the zoom window (Fig. 6p), it exhibits considerably richer details based on the Landsat features while maintaining a reasonable value range compared to the MODIS LST (Fig. 6m). In general, the prediction of ESTARFM on a training date is identical to the fine-resolution scene, whereas the prediction of ubESTARFM on a training date borrows the spatial details of fine-resolution data and has a comparable value range with the coarse-

resolution data.

4.3. Evaluation of fused LST against in-situ LST

Fig. 7 shows the scatterplots of ESTARFM and ubESTARFM LST against in-situ LST at the 12 sites between 2013 and 2021 excluding the training dates. The ubESTARFM LST coincides better with the in-situ LST than the ESTARFM LST at most sites. At some sites, such as Boyagin, Great Western Woodlands, and Ridgefield (Fig. 7a, e, and f, respectively), where the ESTARFM LST is highly biased, the ubESTARFM can adjust the LST values to a more reasonable scale with the in-situ LST. Even at sites where the ESTARFM LST shows better bias performance, such as the Cumberland Plain and Wombat State Forest sites (Fig. 7c and k, respectively), the ubESTARFM LST exhibits a better

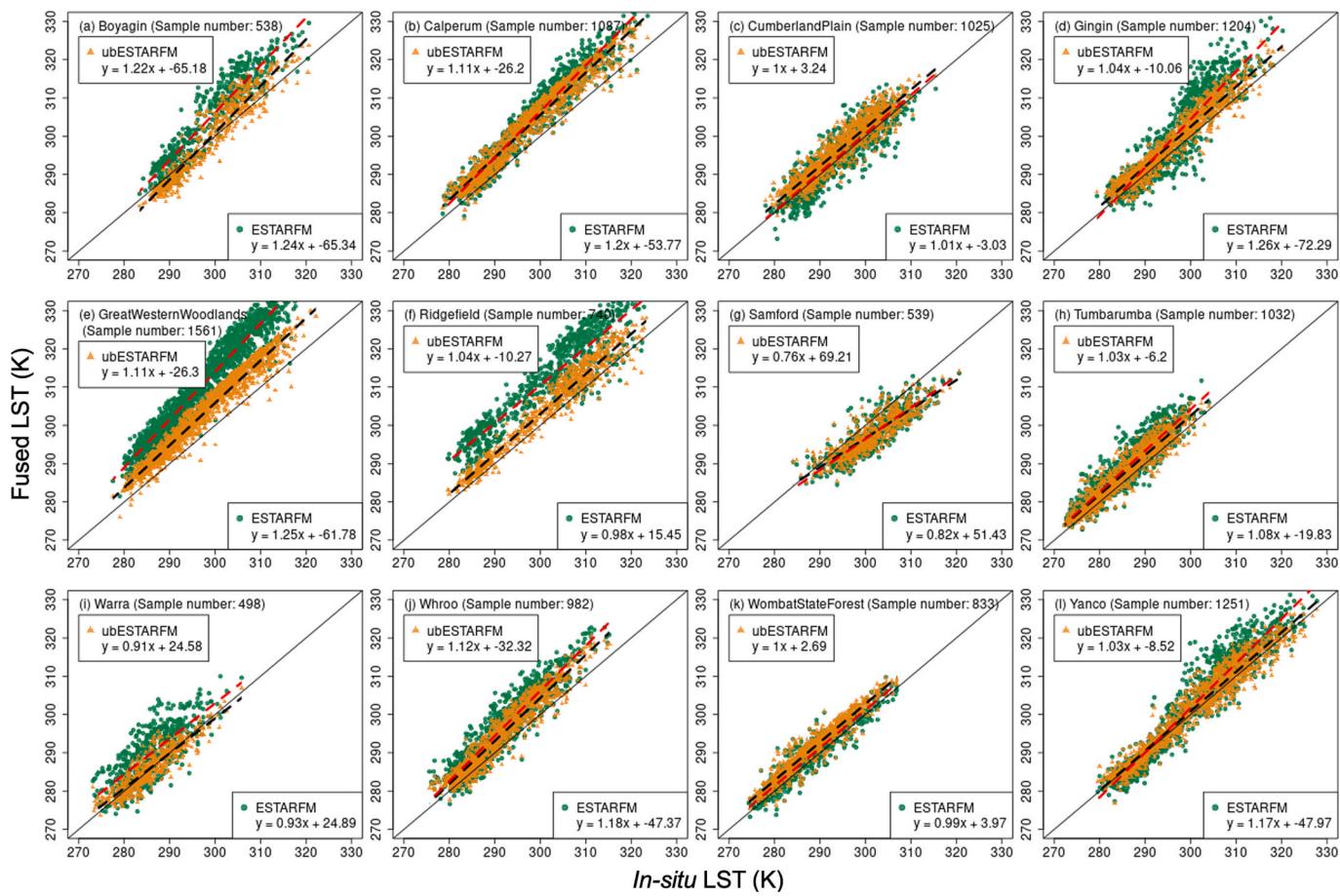


Fig. 7. Scatterplots of ESTARFM and ubESTARFM LST against in-situ LST at the 12 sites (excluding the training dates). The thin black continuous diagonal line on each plot is the 1:1 line. The red and black dashed lines are the linear regression lines of best-fit (as provided by the equations on each sub-part) for ESTARFM and ubESTARFM, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6

Statistics of MODIS, ESTARFM and ubESTARFM LST against in-situ LST across the 12 sites and as grouped into climate classifications (semi-arid, humid subtropical, oceanic and cool temperate) and land cover classes (vegetated and non-vegetated). The training dates are excluded. The units of the Bias and ubRMSE statistics are K.

Site name	Bias			ubRMSE			R			Sample number
	MODIS	ESTARFM	ubESTARFM	MODIS	ESTARFM	ubESTARFM	MODIS	ESTARFM	ubESTARFM	
Boyagin	-0.44	5.90	0.64	3.16	4.28	3.62	0.96	0.94	0.95	538
Calperum	5.19	5.97	5.27	2.92	3.60	2.79	0.97	0.97	0.97	1087
Cumberland Plain	2.81	0.59	2.19	2.75	3.46	2.58	0.93	0.90	0.94	1025
Gingin	2.36	3.20	2.29	2.44	4.77	2.24	0.96	0.92	0.97	1204
Great Western Woodlands	4.82	13.26	5.53	2.50	3.96	2.47	0.98	0.97	0.98	1561
Ridgefield	3.00	10.46	3.03	2.89	4.56	2.58	0.97	0.92	0.98	740
Samford	-2.16	-3.82	-3.74	3.11	3.19	3.03	0.87	0.86	0.87	539
Tumbarumba	1.68	3.04	2.12	2.48	3.10	2.29	0.94	0.92	0.95	1032
Warra	-0.79	3.97	0.21	2.68	4.65	2.49	0.90	0.76	0.91	498
Whroo	3.71	4.66	3.28	3.18	3.83	2.75	0.96	0.93	0.96	982
Wombat State Forest	1.94	1.30	2.77	1.97	2.18	1.96	0.96	0.96	0.97	833
Yanco	0.77	1.74	0.82	2.80	4.05	2.66	0.97	0.96	0.97	1251
Semi-arid	3.13	5.78	3.44	2.55	3.83	2.45	0.97	0.96	0.97	5936
Humid subtropical	-1.30	1.03	-1.55	3.14	3.74	3.32	0.91	0.90	0.91	1077
Oceanic	2.23	6.14	2.50	2.66	3.71	2.41	0.95	0.92	0.96	1772
Cool temperate	2.45	2.86	2.23	2.90	3.84	2.63	0.94	0.89	0.94	2505
Vegetated	1.78	3.26	1.81	2.70	3.76	2.54	0.95	0.91	0.95	7391
Non-vegetated	3.62	7.53	3.94	2.71	3.89	2.62	0.97	0.97	0.98	3899
All	2.42	4.73	2.55	2.70	3.80	2.57	0.95	0.92	0.95	11,290

performance of ubRMSE, implying less deviation in error in the evaluation against in-situ LST.

Table 6 summarises the performance statistics of the MODIS, ESTARFM, and ubESTARFM LST against in-situ LST at the 12 sites, excluding the training dates. The results show that the ubESTARFM LST

outperforms the ESTARFM LST in terms of bias at 10 out of 12 sites, except for the Cumberland Plain and Wombat State Forest sites. The overall bias performance of the ubESTARFM LST is 2.55 K, which is about half of that of the ESTARFM LST (4.73 K) and comparable with the MODIS LST (2.42 K) benchmark. The ubESTARFM LST also has a better

performance of ubRMSE against in-situ LST than both the MODIS and ESTARFM LST at 11 out of 12 sites, except for Boyagin. The overall ubRMSE performance of the ubESTARFM LST is 2.57 K, whereas that of the ESTARFM and MODIS LST are 3.80 and 2.70 K, respectively. In terms of R, the performance of ubESTARFM LST is better than that of ESTARFM LST at all sites and is very similar to MODIS LST's performance. The overall R performance of the ubESTARFM, ESTARFM, and MODIS LST are 0.95, 0.92, and 0.95, respectively. When it comes to the varying climate and vegetation groups, ubESTARFM has a better performance of all three metrics than ESTARFM under most conditions, except for the humid subtropical group, where ubESTARFM and ESTARFM have bias of -1.55 K and 1.03 K, respectively. These metrics, which are based on 11,290 samples, indicate that the ubESTARFM algorithm not only eliminates potentially systematic biases in the fusion process but also provides a better coincidence with the in-situ LST compared to the ESTARFM algorithm. The evaluations across various climates and vegetation conditions showcase the robustness of ubESTARFM. Furthermore, compared to MODIS LST, ubESTARFM LST can reduce the deviation of uncertainty and maintain the same level of correlation with in-situ LST.

Fig. 8 presents the boxplots of seasonal statistics of MODIS, ESTARFM and ubESTARFM LST against in-situ LST across the 12 sites. The metrics of ubESTARFM LST consistently outperform those of ESTARFM LST in all seasons and are comparable with those of MODIS LST, with the ubRMSE of ubESTARFM LST even slightly better than that of MODIS LST. In summer, all three LSTs exhibit relatively higher bias

and ubRMSE than other seasons, while in winter, all three LSTs demonstrate the best performance in terms of bias and ubRMSE. The gaps between ESTARFM LST and ubESTARFM LST with respect to bias and ubRMSE are also the narrowest in the winter season. These results suggest that the typical ESTARFM algorithm may have more uncertainty when the land surface is heated up. Moreover, with regard to the metric of R, the three LSTs demonstrate relatively similar performance in the spring and autumn seasons, while in the summer and winter seasons, the median values of ESTARFM LST show gaps of approximately 0.15 compared to the other two LSTs.

Fig. 9 shows the temporal variations of MODIS, ESTARFM and ubESTARFM LST at the 12 sites between 01/Jan/2013 and 31/Dec/2021. All three LSTs have reasonable seasonal variations that are consistent with the in-situ LST. For the sites where the ESTARFM LST is remarkably biased (Boyagin, Great Western Woodlands, Ridgefield and Warra), the ubESTARFM LST is able to narrow the gap with the in-situ LST hence provides better temporal consistency. Furthermore, when the MODIS LST is discrete, e.g., summer in 2017 and 2018 at the Cumberland Plain site (Fig. 9c), the ubESTARFM LST shows less uncertainty and better matches the trend of in-situ LST. At the Warra site (located in Tasmania and has an annual rainfall of around 1700 mm), both MODIS and ESTARFM LST have relatively high uncertainty, while the ubESTARFM can still have a stable performance. Fig. S1 gives shortened LST time series of the 12 sites within a single year to show more details in temporal variations of LST.

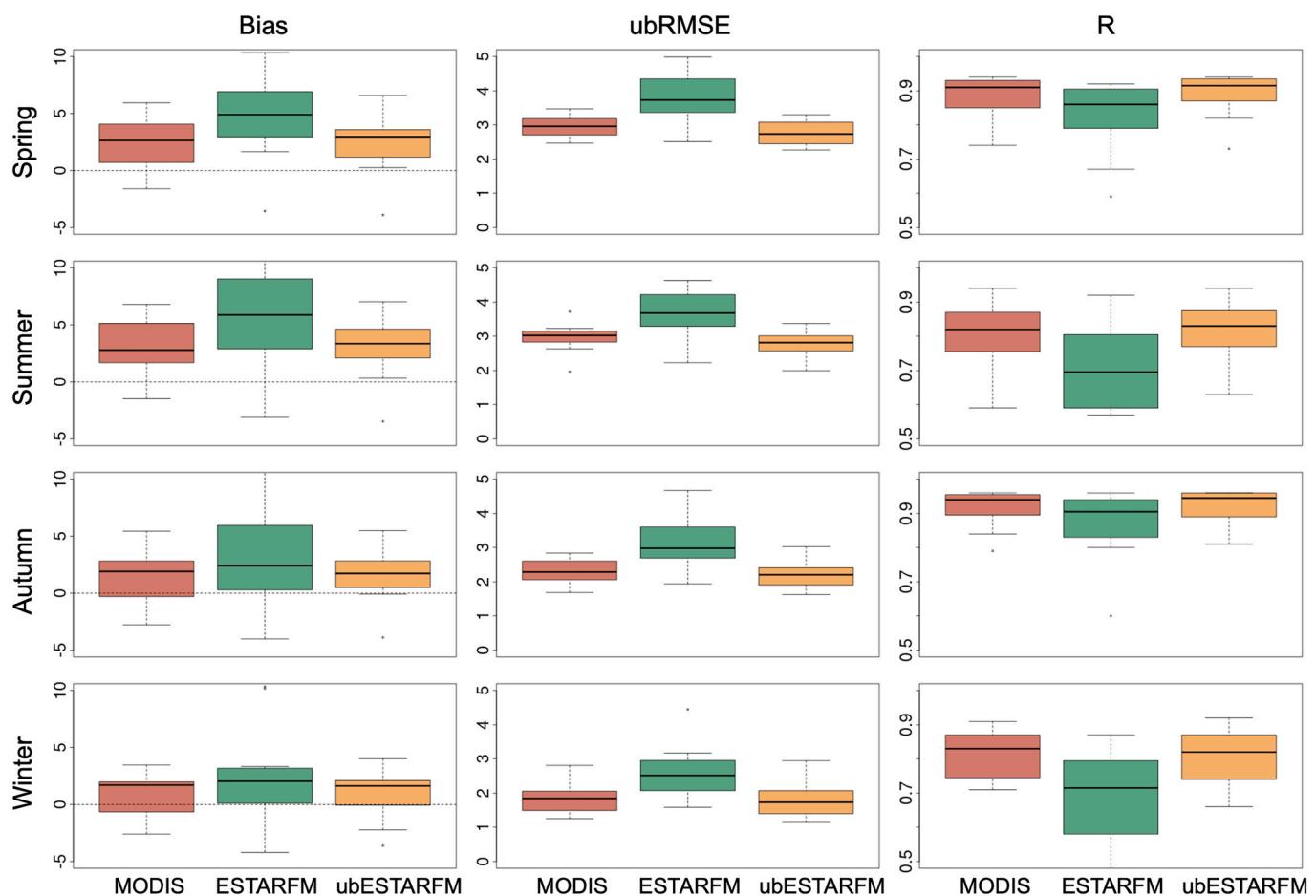


Fig. 8. Seasonal boxplots of statistics of MODIS, ESTARFM and ubESTARFM LST against in-situ LST at the 12 sites (excluding the training dates). For the southern hemisphere temperate zone, spring is from September to November, summer from December to February, autumn from March to May, and winter from June to August. The dark horizontal line within the boxplots represents the median value; the upper and lower quartiles of the interquartile range represent the 75% and 25% thresholds, respectively; the upper and lower whiskers extend from the box represent the maximum and minimum values excluding any outliers. The units of the Bias and ubRMSE statistics are K.

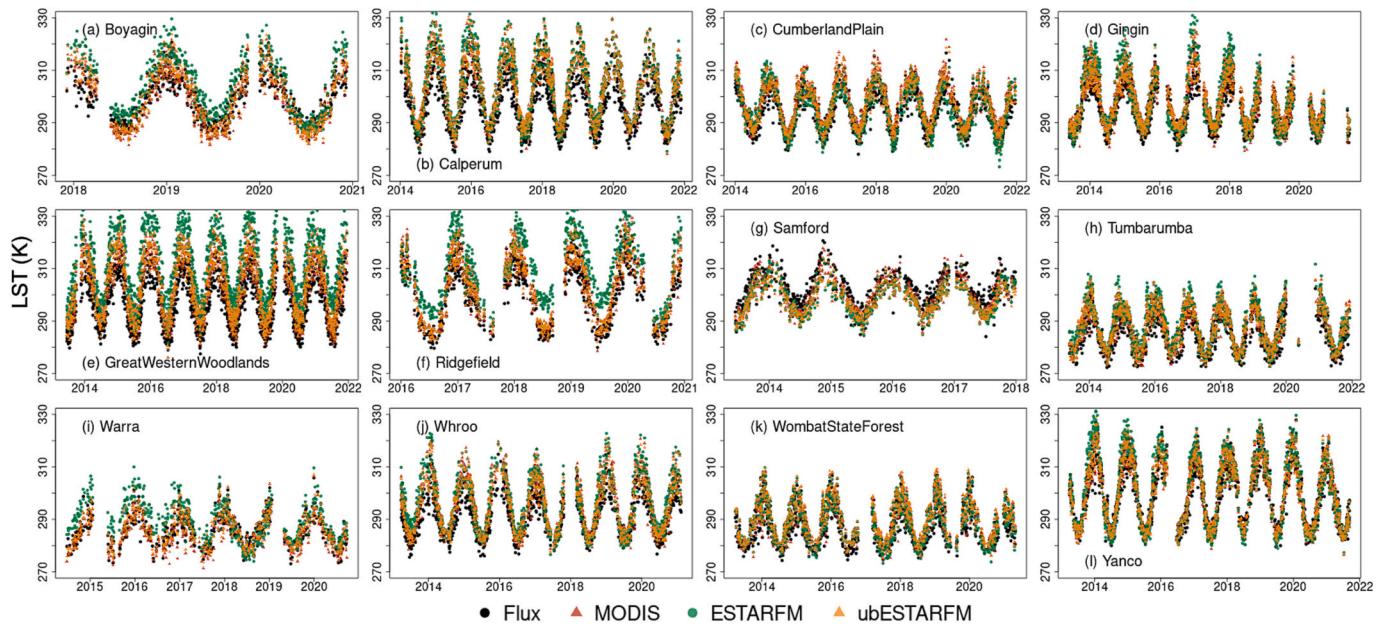


Fig. 9. Temporal variations of MODIS, ESTARFM and ubESTARFM LST at the 12 sites between 2013 and 2021. Gaps in the time series are either due to the absence of in-situ LST or cloud contamination in the MODIS LST. A shortened time series for only 1 year can be found in Fig. S1.

4.4. Spatial comparison between fused LST and ECOSTRESS LST

We selected two examples to show the spatial difference between the MODIS, ECOSTRESS, ESTARFM and ubESTARFM LST on clear sky days (Fig. 10). At the Great Western Woodlands site (Fig. 10a-h), the ESTARFM LST is positively biased and has spatial discontinuity (Fig. 10c). The MODIS, ECOSTRESS and ubESTARFM LST have a consistent spatial pattern at the ROI scale (Fig. 10a, b, d), with the ECOSTRESS LST being the sharpest, likely due to its native resolution of 70 m, which is even finer than the Landsat LST. In the zoom windows, the ESTARFM LST provides a correct spatial pattern with some distinct paths but with approximately 7 K higher values (Fig. 10g); whereas the ubESTARFM LST shows both reasonable values and abundant spatial details (Fig. 10h). At the Tumbarumba site (Fig. 10i-p), all four LSTs show a consistent spatial pattern at the ROI scale (Fig. 10i-l), though they have some differences in values at the southwest region. However, in the zoom windows, both the values and spatial pattern of ESTARFM LST (Fig. 10o) are considerably different with both MODIS (Fig. 10m) and ECOSTRESS LST (Fig. 10n). As a comparison, the ubESTARFM LST still performs well in both numerical values and spatial structure (Fig. 10p).

Table 7 provides a summary of the pixel-wise statistics of ESTARFM and ubESTARFM LST against ECOSTRESS LST. Compared to the ESTARFM LST, the ubESTARFM LST has a lower bias against ECOSTRESS LST at 7 sites, and a slightly better overall performance (-1.69 K to 1.79 K). However, the ubESTARFM LST is remarkably superior in terms of ubRMSE and R against ECOSTRESS LST, where it outperforms the ESTARFM LST at 11 out of 12 sites (except for Cumberland Plain). The overall ubRMSE of ESTARFM and ubESTARFM LSTs is 2.68 and 2.00 K , respectively, while the overall R of them is 0.59 and 0.70 , respectively. In general, the cross-satellite comparison demonstrated that ubESTARFM has a comparable performance of bias with ESTARFM, but shows a substantial improvement in uncertainty reduction and correlation with independent satellite LST retrievals.

5. Discussion

5.1. A further evaluation of ubESTARFM's assumption

5.1.1. Comparison between different bias correction strategies

This research demonstrated the superiority of ubESTARFM over ESTARFM on a continental scale, considering that MODIS LST generally exhibited better performance across most sites. However, the evaluation of ubESTARFM's effectiveness was not fully explored, as its applicability is expected to be evident when coarse-resolution LST data presents comparable or even slightly degraded performance compared to fine-resolution LST data. Therefore, to further assess the applicability of ubESTARFM, we selected six sites where Landsat LST demonstrated comparatively superior agreements with in-situ LST in comparison to MODIS LST (as indicated in Fig. 4), including Cumberland Plain, Gingin, Tumbarumba, Whroo, Wombat State Forest, and Yanco. Moreover, we incorporated three additional methods as comparative benchmarks: (1) an alternative bias correction strategy utilising fine-resolution data as a reference (described by Eq. (12)), referred to as ubESTARFM_{fine}; (2) a temporal bias correction approach proposed by Long et al. (2020); and (3) a statistical downscaling approach for sharpening thermal imagery (TsHARP; Agam et al., 2007).

Fig. 11 are boxplots of the bias, ubRMSE, and R values for the five methods evaluated at three sites (where Landsat is better) and six sites (where Landsat is better or comparable), respectively. The statistics for all the 12 sites are summarised in Table S3. The ubESTARFM approach consistently exhibits the best performance in terms of bias, ubRMSE, and R at the six sites, with mean values of 2.17 K , 2.43 K , and 0.96 , respectively. Additionally, ubESTARFM achieves the best mean ubRMSE (2.45 K) and R (0.96) at the three sites, while its mean bias of 1.79 K is second only to ESTARFM (1.25 K). These results demonstrate the robustness of ubESTARFM to produce results that both have high accuracy while maintaining a low level of uncertainty, marking a substantial improvement over ESTARFM, which yields mean ubRMSE values of 3.36 K for the three sites and 3.66 K for the six sites. Comparatively, the ubESTARFM_{fine} approach exhibits only a minor degradation in performance compared to ubESTARFM, which can be attributed to the residual uncertainty retained from the fine-resolution LST. Furthermore, Long's method demonstrates improved mean

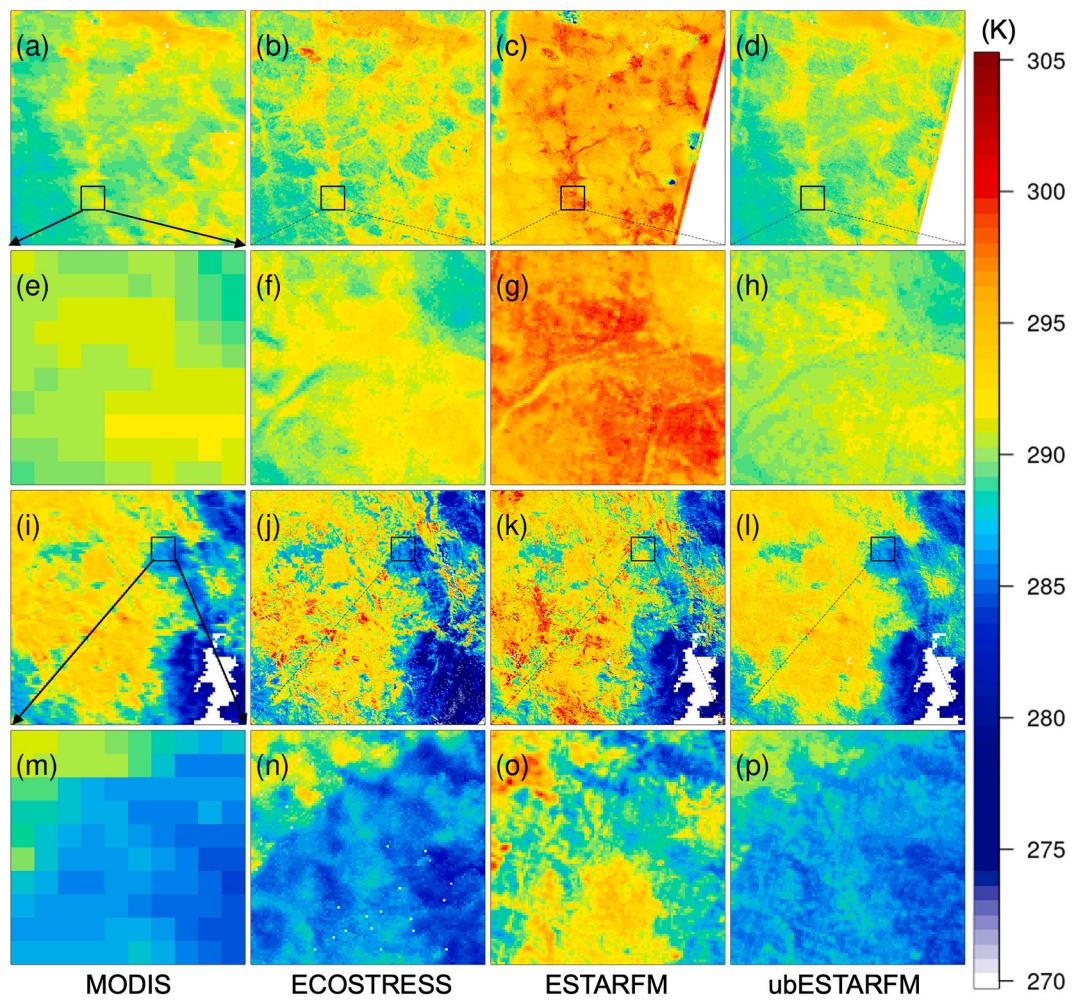


Fig. 10. Two examples of the spatial comparison between MODIS, ECOSTRESS, ESTARFM and ubESTARFM LST on clear sky days. (a-d) are the LSTs within a homogenous $1^\circ \times 1^\circ$ ROI (Great Western Woodlands) on 11/Jul/2019, and the black window is a $10\text{ km} \times 10\text{ km}$ area surrounding the in-situ site; (e-h) are zoomed areas of the window in (a-d); (i-l) are the LSTs within a heterogenous ROI (Tumbarumba) on 10/Sep/2020, with the black window representing the surrounding area of the in-situ site; (m-p) are zoomed areas of the window in (i-l).

Table 7

Statistics of ESTARFM and ubESTARFM LST against ECOSTRESS LST. The units of the Bias and ubRMSE statistics are K. For the scene number column only ECOSTRESS scenes with >50% clear-sky pixels within the ROIs were considered in the comparison. For the Warra site, N/A means 'Not Applicable', and this is because there was no near clear-sky ECOSTRESS scene having a coincident overpass time with the MODIS LST for the Warra ROI between 01/Jul/2018 and 01/Dec/2021.

Site name	Mean bias		Mean ubRMSE		Mean R		Scene number
	ESTARFM	ubESTARFM	ESTARFM	ubESTARFM	ESTARFM	ubESTARFM	
Boyagin	5.70	-2.62	2.98	1.88	0.54	0.64	8
Calperum	-0.43	-0.04	3.27	2.46	0.38	0.57	4
Cumberland Plain	-1.28	-3.53	2.25	2.27	0.86	0.86	1
Gingin	-0.15	-1.40	2.93	2.63	0.69	0.74	6
Great Western Woodlands	2.78	-1.89	1.97	1.02	0.53	0.83	3
Ridgefield	5.70	-2.62	2.98	1.88	0.54	0.64	8
Samford	-2.68	-1.89	2.87	2.68	0.61	0.64	1
Tumbarumba	-0.41	-1.24	2.40	2.20	0.88	0.90	2
Warra	N/A	N/A	N/A	N/A	N/A	N/A	0
Whroo	0.41	0.45	2.26	1.47	0.36	0.58	2
Wombat State Forest	-2.94	-2.16	2.28	2.01	0.76	0.85	4
Yanco	-1.11	-0.20	1.90	1.77	0.64	0.66	4
All	1.79	-1.69	2.68	2.00	0.59	0.70	43

ubRMSE values compared to ESTARFM, specifically 3.02 K for the three sites and 3.22 K for the six sites. This highlights the benefits of the temporal bias correction approach, which relies on the uncertainty level of a single dataset rather than directly accumulating uncertainties. However, Long's method displays unstable bias performance, likely due

to the unsuitability of using Landsat LST as the temporal reference. The remarkably sparser temporal frequency of Landsat (i.e., 1/16 of MODIS observations) renders the characteristics of Landsat LST time-series less representative for generating daily results. Regarding TsHARP LST, its performance consistently falls slightly behind that offered by

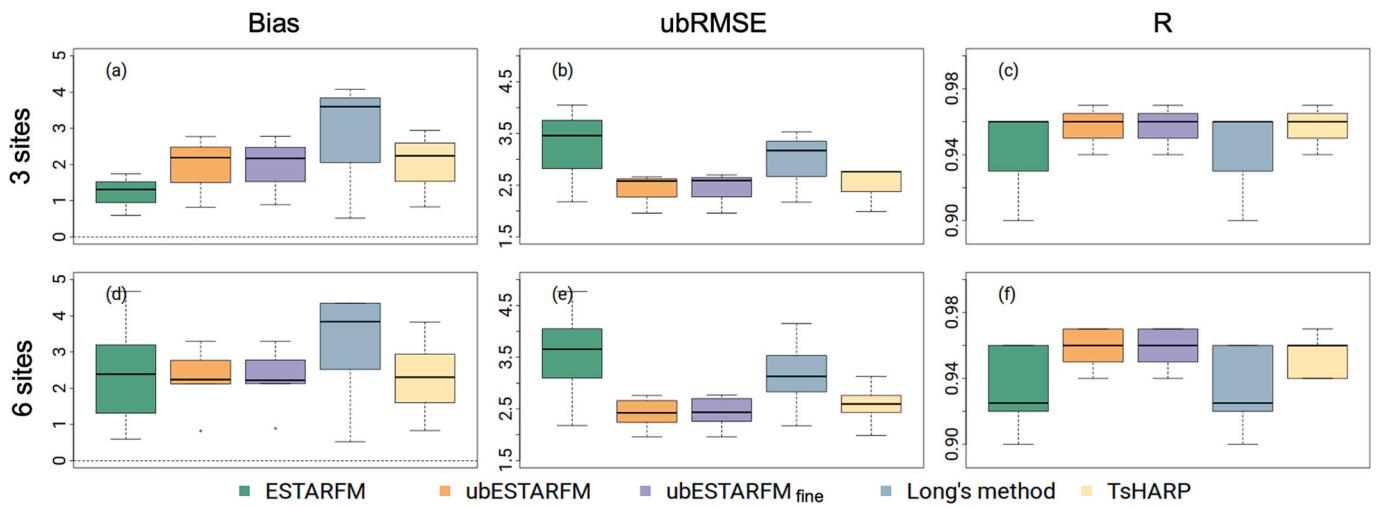


Fig. 11. Boxplots of statistics of ESTARFM, ubESTARFM, ubESTARFM fine, Long's method and TsHARP LST against in-situ LST (excluding the training dates). Parts (a-c) present the statistics from 3 sites (Cumberland Plain, Wombat State Forest and Yanco) where Landsat LST has better agreement against in-situ LST than MODIS LST. Parts (d-f) present the statistics from 6 sites (Cumberland Plain, Gingin, Tumbarumba, Whroo, Wombat State Forest and Yanco) where Landsat LST has better or comparable agreement against in-situ LST compared to MODIS LST. The explanation of boxplot features is the same as in Fig. 8 caption. For the boxplots in (a-c), the upper and lower quartiles for the interquartile range lack statistical significance due to the limited data available from only 3 sites. The sample number for 3 sites and 6 sites are 3104 and 6273, respectively. The units of the Bias and ubRMSE statistics are K.

ubESTARFM to estimate LST. Nonetheless, the downscaling strategy employed by TsHARP partly incorporates the coarse-resolution residual of MODIS LST to its fine resolution prediction (Agam et al., 2007; Huryna et al., 2019; Lacerda et al., 2021), which is likely to weaken the sharpening effects. To summarise, ubESTARFM was effective when the fine-resolution LST data performs slightly better or shows comparable performance to the coarse-resolution LST data. It is anticipated that ubESTARFM can maintain its effectiveness as long as the uncertainty of coarse-resolution LST does not exceed a difference of 1–1.5 K compared to the fine-resolution LST.

5.1.2. Application of ubESTARFM using geostationary LST

The ubESTARFM algorithm's efficacy in reducing error deviation within the context of MODIS-Landsat LST fusion has been conclusively demonstrated. However, as there has been a notable upsurge in the availability of LST data in recent years, it becomes imperative to assess the versatility of the ubESTARFM algorithm to be applied with LST from diverse sources. Geostationary satellites exhibit the capability to monitor sub-hourly thermal variations over specific regions while with a relatively coarser resolution compared to polar-orbiting satellites (Freitas et al., 2013). A pertinent example is the Himawari-8 mission (Bessho et al., 2016), which furnishes data with a spatial resolution of 2 km and a temporal frequency of 10 min over the Asia-Pacific region, allowing for synchronisation with the overpass time of both MODIS and Landsat satellites. Against this backdrop, we proceeded to apply ubESTARFM using a Himawari-8 LST dataset from the Australian National University (Yu et al., 2024) to further evaluate its effectiveness.

We selected three sites over southeast Australia, including the Tumbarumba, Wombat State Forest and Yanco sites, for the assessment (Fig. 12). The ubESTARFM LST demonstrates ubRMSE (R) values of 1.81 K (0.97), 2.29 K (0.96) and 2.57 K (0.98) for the respective sites. These values, when compared to those generated by ESTARFM, being 3.72 K (0.89), 4.20 K (0.87) and 4.31 K (0.94), respectively, demonstrate a substantial improvement. Notably, the performance of ubESTARFM LST also surpasses that of Himawari-8 LST (1.84 K and 2.89 K) at the Tumbarumba and Yanco sites, respectively. In terms of spatial patterns, similar to the case of MODIS-Landsat fusion, while ESTARFM LST captures intricate spatial features, it struggles to accurately estimate value ranges, with observed overestimations among all three sites (Fig. 12 g-i). In contrast, ubESTARFM LST not only enhances spatial details but also

provides more plausible numeric value estimations (Fig. 12 j-l). These outcomes underscore the potential of the ubESTARFM algorithm for application using LST data acquired from geostationary platforms.

5.2. Strengths and prospects of ubESTARFM

Spatiotemporal fusion algorithms have been increasingly used to generate fine-resolution LST data for agricultural fields and other studies requiring access to fine-resolution LST data. However, since LST data are sensitive to time-of-day of image acquisition and may have retrieval process uncertainties, original spatiotemporal fusion approaches (e.g., STARFM and ESTARFM) may generate LST data with sub-optimal accuracies (see Table 1). This motivated us to develop and evaluate a fusion variant. Based on the usage of independent ground-based observation for validation and satellite-based evaluation, our results demonstrate that ubESTARFM performs better than ESTARFM in terms of bias control, reduction of error deviation, and maintaining correlation. The strengths and prospects of the ubESTARFM algorithm can be summarized as follows:

The ubESTARFM approach is effective in avoiding the accumulation of systematic errors from multiple sensors. While the difference in overpass time of Landsat and MODIS satellites may not substantially impact reflectance studies, this time difference does have a relatively higher impact in LST studies. LST data derived using ESTARFM are highly reliant on the fine-resolution LST data. In contrast, ubESTARFM calibrates systematic bias of fine-resolution LST using the values of coarse-resolution data within the moving window as a reference while preserving the spatial heterogeneity of fine-resolution data. It's worth noting that the ubESTARFM approach has been shown to considerably reduce the degree of error dispersion (represented by ubRMSE) when compared to in-situ LST measurements, even outperforming the MODIS LST benchmark. This demonstrates that the spatial details and temporal variations of fine features in Landsat LST data can be considered reliable, despite being subject to coherent systematic biases (Duan et al., 2021; Li et al., 2021a; Vlassova et al., 2014). Using ubESTARFM enabled us to generate highly accurate LST data, which is crucial for applications such as modelling fine-resolution ET and soil moisture that are require accurate values of LST (Long et al., 2019; Long and Singh, 2012b) and urban heat island monitoring (Liu et al., 2016; Xia et al., 2019).

Based on the pattern of generated 100 m LST, it appears that

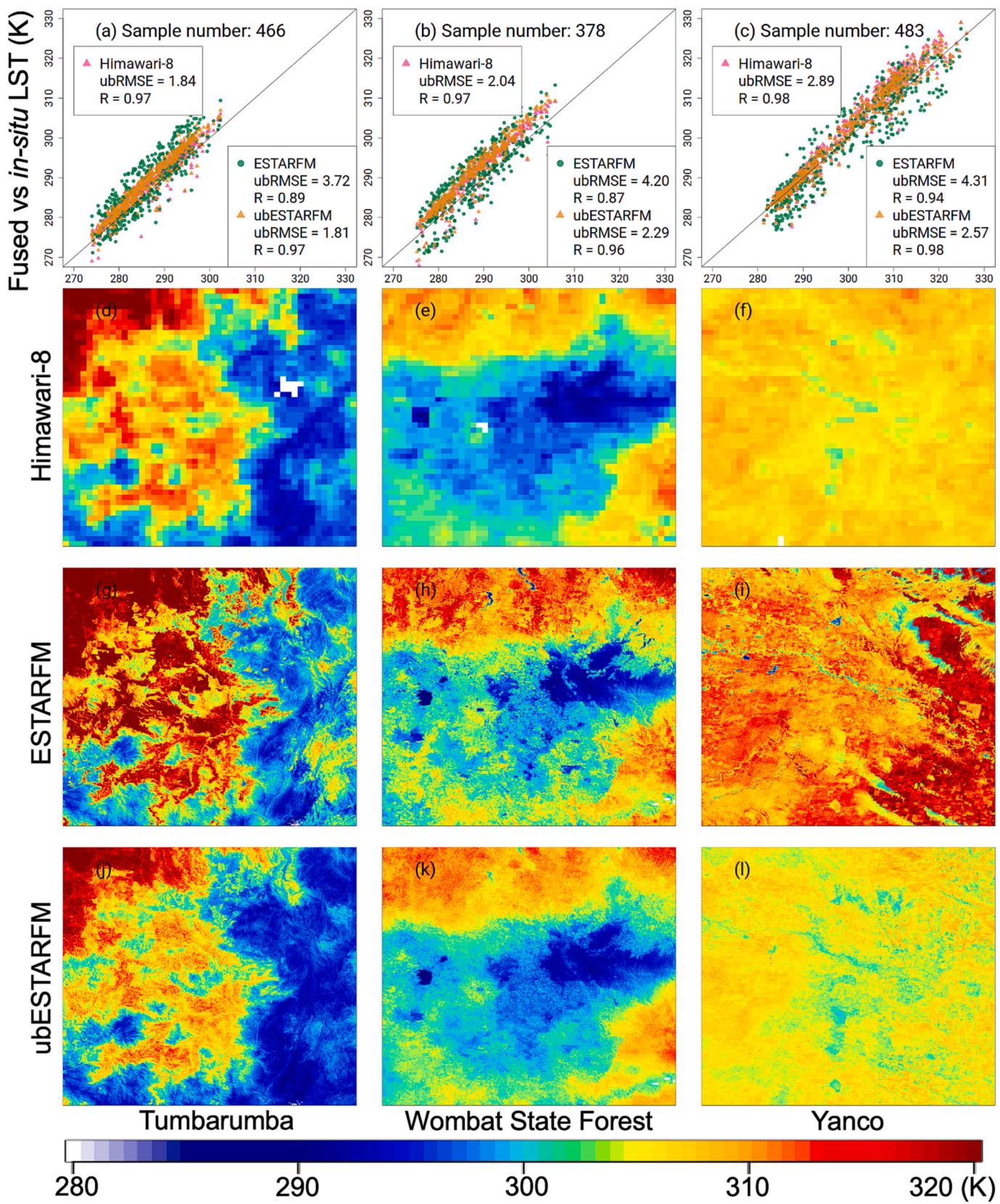


Fig. 12. (a-c) Scatterplots of Himawari-8, ESTARFM and ubESTARFM LST against in-situ LST at the Tumbarumba, Wombat State Forest and Yanco sites, respectively, between 01/Jan/2016 and 31/Dec/2021 (excluding the training dates). The thin black continuous diagonal line on each plot is the 1:1 line. (d-f) are spatial examples of Himawari-8 LST at the three sites on 24/Dec/2018, 18/Nov/2019 and 11/Mar/2020, respectively; (g-i) are spatial examples of ESTARFM LST at the three sites on corresponding dates, respectively; and (j-l) are spatial examples of ubESTARFM LST at the three sites on corresponding dates, respectively. The sample numbers of Himawari-8 LST are less than those of MODIS LST because Himawari-8 LST is only available since July 2015.

ubESTARFM functions more like a downscaling method rather than a traditional fusion approach (Chen et al., 2023; Pu and Bonafoni, 2023). This is because ubESTARFM incorporates both the numerical values from coarse-resolution data and the spatial structure from fine-resolution data, resulting in a simple and robust approach with high computational efficiency. However, for this approach to work optimally, it is necessary for the coarse-resolution data to be at least comparable with the fine resolution data, which may not always be the case. For example, herein Landsat LST performed better than the MODIS LST at 3 of the 12 sites (Table 4). In small spatial extent studies focused on regions where Landsat data are more consistent with in-situ data and where ground stations are absent, the ubESTARFM approach might ignore the value of Landsat data. To address this issue, an optimisation strategy such as the Markov Chain Monte Carlo method (MCMC) (Brooks, 1998) or an error variance comparison method such as the triple collocation approach (McColl et al., 2014) could be considered. For example, by incorporating MCMC, it is theoretically possible to generate a posterior distribution of the predicted fine-resolution LST values, which captures the inherent uncertainty associated with the predictions. This enhanced uncertainty estimation throughout an iterative sampling process provided by MCMC may allow for a more robust characterisation of the prediction uncertainty in ubESTARFM. These approaches could better evaluate the systematic and random errors in multi-sourced data, thereby enabling the determination of the weights of each input data type in the fusion process, rather than simply using the coarse-resolution data as the reference.

Depending on the rates of change of LST signatures being fused, the length of the temporal gap between the two training dates could impact on results. To ensure the most accurate results possible, we searched for Landsat and MODIS scenes with abundant clear-sky pixels. However, this sometimes can lead to a considerable time gap (e.g., several weeks to months) between the two training dates which could result in additional errors in the fused LST when there is a substantial difference between the LST on the target date and the LST on the training dates. This is particularly relevant in regions where there are agricultural activities (e.g., harvesting and irrigation), as land cover types may change rapidly during the growing period (Van Niel et al., 2003; Van Niel and McVicar, 2004), potentially introducing errors in the fused LST. For example, at the Yanco site, the Landsat LST performs better than the MODIS LST (Fig. 4I), possibly because the former is better able to detect transitions between soil and vegetation associated with agricultural activity. To further improve the ubESTARFM approach, additional efforts could be considered, such as the application of a spectral linear theory to help predict potential changes in land cover types. This could help to minimize errors introduced by changes in land cover types during the temporal gap between the training dates (Xia et al., 2019; Zhu et al., 2016).

There are at least three limitations of the current experimental design. Firstly, some parameters could be optimised, such as the size of the patch and moving window, as well as the number of land covers. Following Zhu et al. (2010), we used their default settings where the window is 51×51 pixels and the patch is 200×200 pixels, which are considered suitable for a scene with 1000×1000 pixels. For studies that have a different spatial extent, an iterative or optimisation strategy might be considered to determine the window size and ensure good consistency between the results and the source data distribution. Secondly, the current algorithm assumes that there are a pre-determined number of land covers, which may not be the case in real-world situations. By incorporating a land cover map in the window searching strategy (Liu et al., 2016; Shi et al., 2022), the algorithm may better account for the spatial variability of land cover types and their impact on surface temperature, thus has the potential to generate more accurate and robust fusion results. Nonetheless, this strategy also creates another dependency and reliance on land-cover mapping being accurate and up-to-date, which requires additional efforts and inputs. Thirdly, and finally, the emissivity values used herein were obtained from the UW

Baseline Fit Emissivity Database (Seemann et al., 2008) coupled with ground-based radiation to generate in-situ LST. However, the UW data has a 5 km resolution derived using the same input as the MODIS operational emissivity product. The 5 km resolution is closer to the MODIS spatial resolution and may be sub-optimal for the much finer Landsat product due to the dependency of LST product on emissivity (Mohamed et al., 2017; Parastatidis et al., 2017). Additionally, as the MODIS LST and in-situ LST share an emissivity input, this could also compromise the Landsat LST in the in-situ evaluations. In future studies, we could consider using the emissivity from next-generation products (e.g., ECOSTRESS) to generate in-situ LSTs, which would be more appropriate in fine-resolution studies.

5.3. Future applications

The ubESTARFM algorithm demonstrates an advancement to the typical ESTARFM in generating fine resolution LSTs. There are a few application directions and implications derived from this study:

In this study, the majority of the in-situ sites in this study were covered by croplands, woodlands, or forests. To further improve the algorithm's applicability, future studies should extend to areas covered by sparse vegetation, bare soil, or desert. The Landsat LST appears to be highly biased in semi-arid regions (Table 4), as evidenced by the results at the Great Western Woodlands and Ridgefield sites. Furthermore, the examination of ubESTARFM's performance in urban areas is an aspect that deserves exploration, since urban environments present unique challenges due to the increased complexity and heterogeneity of land cover types, making the systematic differences between MODIS and Landsat data more prominent. Particularly, an accurate estimation of emissivity values can be challenging in urban areas and is often hindered by limitations associated with the commonly used NDVI threshold method (Weng et al., 2014). To shed light on the applicability of ubESTARFM in urban settings, it is important to expand the analysis to include urban areas in addition to the peri-urban regions investigated herein (i.e., Cumberland Plain and Samford). By considering diverse urban landscapes with varying levels of impervious surfaces, vegetation cover, infrastructure and shadowing, a more comprehensive understanding of ubESTARFM's performance can be obtained. Therefore, additional evaluation should be conducted in semi-arid areas with sparse vegetation, such as agricultural (both irrigated and non-irrigated) and horticultural (tree and vine) crops and shrubs, and to further test ubESTARFM's performance in a wider range of land cover types especially in urban environments.

We utilised Landsat 8 and MODIS data to generate LSTs between 01/Jan/2013 and 31/Dec/2021. Our current LSTs were generated for the morning time only, around 11 am local solar time. To better understand the diurnal patterns of LST, future research could include LST at other times by utilising data with higher temporal resolutions, such as the geostationary Himawari-8 AHI LST with a 10-min image acquisition frequency (Yamamoto et al., 2018; Yu et al., 2024) and/or the ECOSTRESS LST which has irregular overpass times (Fisher et al., 2020). To improve spatiotemporal completeness, future applications could incorporate a gap-filled coarse-resolution satellite LST product or LST variables from land surface models, while additional systematic errors must be quantified.

The fine-resolution LST generated herein has broad applications in a variety of fields, such as urban heat island studies, field-to-farm scale ET estimation, and soil moisture retrieval. By coupling the generated LSTs with physical models such as radiative transfer models (Jackson, 1993; Vergopolan et al., 2020), energy balance models (Guerszman et al., 2022; Long and Singh, 2012a; McVicar and Jupp, 1999), or soil-vegetation temperature decomposition algorithms (Yang et al., 2015), we can derive corresponding fine-resolution variables and develop improved farm- or regional-scale water management strategies and decision-making processes.

6. Conclusion

Traditional spatiotemporal fusion approaches may not be suitable for direct application in LST studies due to the high temporal dynamics of LST. Additionally, the effectiveness of traditional spatiotemporal fusion approaches on LST data has not been thoroughly evaluated previously at relatively fine-resolution for a wide range of land cover experienced across continental climate gradients. We developed an unbiased variant of ESTARFM, referred to as ubESTARFM, to better accommodate spatiotemporal dynamics to generate LST data. The ubESTARFM utilises a moving window to locally correct the systematic biases of chosen pixels of fine-resolution LST using the mean value of the corresponding coarse-resolution LST pixels as a reference, which was shown to generate LST having consistency with coarse-resolution data while maintaining spatial heterogeneity.

We evaluated the performance of both algorithms through temporal evaluations against in-situ LST and a spatial comparison with the ECOSTRESS LST product across 12 locations scattered across the Australian continent. Over 11,290 samples at 12 sites, the results demonstrate that ubESTARFM outperforms the ESTARFM method, with a bias of 2.55 K, ubRMSE of 2.57 K, and R of 0.95 (compared to 4.73 K, 3.80 K and 0.92, respectively, for ESTARFM). The deviation of uncertainty (represented by ubRMSE) of ubESTARFM LST is even lower than the benchmark of MODIS LST (2.70 K). In the spatial comparison with the ECOSTRESS LST, the ubESTARFM LST has a bias of -1.69 K, ubRMSE of 2.00 K, and R of 0.70 over 43 near clear-sky scenes, while the ESTARFM LST has a bias of 1.79 K, ubRMSE of 2.68 K, and R of 0.59. Ground-based evaluations and satellite-based comparisons demonstrate that ubESTARFM avoids the accumulation of systematic bias that occurs within ESTARFM, considerably reduce the deviation of uncertainty, and maintain a good level of correlation with validation datasets.

Overall, compared to ESTARFM, ubESTARFM makes better use of the reliable numeric values in MODIS LST while borrowing spatial details from Landsat LST. A further assessment also underscores the potential of ubESTARFM for application using LST data acquired from geostationary platforms (e.g., Himawari-8). Further improvements could include considering potential changes in land cover types between training dates, extending the theory to sparsely vegetated areas and urban, and incorporating data from previous Landsat satellites and/or from upcoming next-generation satellites. It is expected that the generated LST can have broad usage and better enable farm- or regional-scale water management strategy or decision making. Furthermore, the LST data used and generated herein, encompassing a comprehensive spatial extent over diverse land covers and climatic conditions, is publicly available for benchmarking future algorithmic refinements.

CRediT authorship contribution statement

Yi Yu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Luigi J. Renzullo:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Tim R. McVicar:** Methodology, Writing – review & editing. **Brendan P. Malone:** Writing – review & editing. **Siyuan Tian:** Writing – review & editing.

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

The ubESTARFM algorithm is written in R and available at <https://doi.org/10.5281/zenodo.8017282>. The collected LST data (i.e., MODIS, Landsat-8 and ECOSTRESS) and LST data generated herein are

available from the CSIRO Data Access Portal (<https://doi.org/10.25919/b77m-8n31>).

Acknowledgements

This research was performed as part of a PhD by the first author (YY) under an academic collaboration between the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). This research was undertaken while supported by the ANU University Research Scholarship and an ANU-CSIRO Digital Agriculture Supplementary Scholarship through the Centre for Entrepreneurial Agri-Technology (CEAT). We thank the continued support of the TERN Landscapes Observatory (<https://www.tern.org.au/tern-observatory/tern-landscapes/>), a sensing platform of the Terrestrial Ecosystem Research Network (TERN; <https://www.tern.org.au/>), which is supported and enabled by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS). We acknowledge the resources and services provided by the National Computational Infrastructure (NCI), which is also supported by the Australian Government through NCRIS. This research was supported by resources and expertise provided by CSIRO IMT Scientific Computing. We wish to thank the RSE Editorial team and the three anonymous reviewers for their constructively critical comments which provided the catalyst to improve an earlier version of this research.

Appendix A. Supplementary data

Supplementary materials to this article can be found online at <https://doi.org/10.1016/j.rse.2023.113784>.

References

- Abowarda, A.S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., Sun, Z., 2021. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. *Remote Sens. Environ.* 255, 112301. <https://doi.org/10.1016/j.rse.2021.112301>.
- Agam, N., Kustas, W.P., Anderson, M.C., Li, F., Neale, C.M., 2007. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sens. Environ.* 107, 545–558. <https://doi.org/10.1016/j.rse.2006.10.006>.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. In: , 300. FAO, Rome, p. D05109.
- Anderson, M., Norman, J., Kustas, W., Houborg, R., Starks, P., Agam, N., 2008. A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. *Remote Sens. Environ.* 112, 4227–4241. <https://doi.org/10.1016/j.rse.2008.07.009>.
- 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 Sens. Environ.* 252, 112189. <https://doi.org/10.1016/j.rse.2020.112189>.
- Barsi, J.A., Schott, J.R., Hook, S.J., Raqueno, N.G., Markham, B.L., Radocinski, R.G., 2014. Landsat-8 thermal infrared sensor (TIRS) vicarious radiometric calibration. *Remote Sens* 6, 11607–11626. <https://doi.org/10.3390/rs61111607>.
- Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci. Data* 5, 1–12. <https://doi.org/10.1038/sdata.2018.214>.
- Beringer, J., Hutley, L.B., McHugh, I., Arndt, S.K., Campbell, D., Cleugh, H.A., Cleverly, J., Resco de Dios, V., Eamus, D., Evans, B., 2016. An introduction to the Australian and New Zealand flux tower network - OzFlux. *Biogeosciences* 13, 5895–5916. <https://doi.org/10.5194/bg-13-5895-2016>.
- Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., Kumagai, Y., Miyakawa, T., Murata, H., Ohno, T., 2016. An introduction to Himawari-8/9 - Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn. Ser. II* 94, 151–183. <https://doi.org/10.2151/jmsj.2016-009>.
- Brooks, S., 1998. Markov chain Monte Carlo method and its application. *J. R. Stat. Soc. Ser. D (The Statistician)* 47, 69–100. <https://doi.org/10.1111/1467-9884.00117>.
- Chen, G., Lu, H., Zou, W., Li, L., Emam, M., Chen, X., Jing, W., Wang, J., Li, C., 2023. Spatiotemporal fusion for spectral remote sensing: a statistical analysis and review. *J. King Saud Univ. Comput. Inform. Sci.* 35, 259–273. <https://doi.org/10.1016/j.jksuci.2023.02.021>.
- Chen, J., Brissette, F.P., Leconte, R., 2011. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J. Hydrol.* 401, 190–202. <https://doi.org/10.1016/j.jhydrol.2011.02.020>.
- Chu, H., Luo, X., Ouyang, Z., Chan, W.S., Dengel, S., Biraud, S.C., Torn, M.S., Metzger, S., Kumar, J., Arain, M.A., 2021. Representativeness of Eddy-covariance flux footprints

- for areas surrounding AmeriFlux sites. Agric. For. Meteorol. 301, 108350. <https://doi.org/10.1016/j.agrformet.2021.108350>.
- Cristóbal, J., Jiménez-Muñoz, J.C., Prakash, A., Mattar, C., Skoković, D., Sobrino, J.A., 2018. An improved single-channel method to retrieve land surface temperature from the Landsat-8 thermal band. Remote Sens 10, 431. <https://doi.org/10.3390/rs10030431>.
- Duan, S.-B., Li, Z.-L., Zhao, W., Wu, P., Huang, C., Han, X.-J., Gao, M., Leng, P., Shang, G., 2021. Validation of Landsat land surface temperature product in the conterminous United States using in situ measurements from SURFRAD, ARM, and NDBC sites. Int. J. Digit. Earth 14, 640–660. <https://doi.org/10.1080/17538947.2020.1862319>.
- 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. Agric. Water Manag. 269, 107664. <https://doi.org/10.1016/j.agwat.2022.107664>.
- Eleftheriou, D., Kiachidis, K., Kalmintzis, G., Kalea, A., Bantasis, C., Koumadoraki, P., Spathara, M.E., Tsolaki, A., Tzampazidou, M.I., Gemitz, A., 2018. Determination of annual and seasonal daytime and nighttime trends of MODIS LST over Greece - climate change implications. Sci. Total Environ. 616, 937–947. <https://doi.org/10.1016/j.scitotenv.2017.10.226>.
- Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., van Dijk, A.I.J.M., 2013. Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: a framework for algorithm selection. Remote Sens. Environ. 133, 193–209. <https://doi.org/10.1016/j.rse.2013.02.007>.
- 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 Resour. Res. 56, e2019WR026058 <https://doi.org/10.1029/2019WR026058>.
- Freitas, S.C., Trigo, I.F., Macedo, J., Barroso, C., Silva, R., Perdigão, R., 2013. Land surface temperature from multiple geostationary satellites. Int. J. Remote Sens. 34, 3051–3068. <https://doi.org/10.1080/01431161.2012.716925>.
- Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44, 2207–2218. <https://doi.org/10.1109/TGRS.2006.872081>.
- Gevaert, C.M., García-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sens. Environ. 156, 34–44. <https://doi.org/10.1016/j.rse.2014.09.012>.
- Guerschman, J.P., McVicar, T.R., Vleeshouwer, J., Van Niel, T.G., Peña-Arancibia, J.L., Chen, Y., 2022. Estimating actual evapotranspiration at field-to-continent scales by calibrating the CMRSET algorithm with MODIS, VIIRS, Landsat and Sentinel-2 data. J. Hydrol. 605, 127318. <https://doi.org/10.1016/j.jhydrol.2021.127318>.
- Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., White, J.C., 2009. A new data fusion model for high spatial-and-temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sens. Environ. 113, 1613–1627. <https://doi.org/10.1016/j.rse.2009.03.007>.
- Hook, S., Hulley, G., 2019. ECOSTRESS Land Surface Temperature and Emissivity Daily L2 Global 70 m V001 [Dataset]. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/ECOSTRESS/ECO2LSTE.001>. Accessed 19/Mar/2023.
- Hu, L., Brunsell, N.A., Monaghan, A.J., Barlage, M., Wilhelmi, 2014. How can we use MODIS land surface temperature to validate long-term urban model simulations? J. Geophys. Res. Atmos. 119, 3185–3201. <https://doi.org/10.1002/2013JD021101>.
- Hu, T., Renzullo, L.J., van Dijk, A.I.J.M., He, J., Tian, S., Xu, Z., Zhou, J., Liu, T., Liu, Q., 2020. Monitoring agricultural drought in Australia using MT-SAT-2 land surface temperature retrievals. Remote Sens. Environ. 236, 111419. <https://doi.org/10.1016/j.rse.2019.111419>.
- Huang, B., Zhang, H., Song, H., Wang, J., Song, C., 2013. Unified fusion of remote-sensing imagery: generating simultaneously high-resolution synthetic spatial-temporal-spectral earth observations. Remote Sens. Lett. 4, 561–569. <https://doi.org/10.1080/2150704X.2013.769283>.
- Hulley, G.C., Götsche, 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. 60, 1–23. <https://doi.org/10.1109/TGRS.2021.3079879>.
- Huryna, H., Cohen, Y., Karnieli, A., Panov, N., Kustas, W.P., Agam, N., 2019. Evaluation of TsHARP utility for thermal sharpening of Sentinel-3 satellite images using Sentinel-2 visual imagery. Remote Sens 11, 2304. <https://doi.org/10.3390/rs11192304>.
- Jackson, T.J., 1993. III. Measuring surface soil moisture using passive microwave remote sensing. Hydrol. Process. 7, 139–152. <https://doi.org/10.1002/hyp.3600070205>.
- Jiménez-Muñoz, J.C., Sobrino, J.A., Skoković, D., Mattar, C., Cristóbal, 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>.
- Justice, C.O., Vermote, E., Townsend, J.R., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A., 1998. The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. IEEE Trans. Geosci. Remote Sens. 36, 1228–1249. <https://doi.org/10.1109/36.701075>.
- Kalma, J.D., McVicar, T.R., McCabe, M.F., 2008. Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data. Surv. Geophys. 29, 421–469. <https://doi.org/10.1007/s10712-008-9037-z>.
- Kerr, Y.H., Lagouarde, J.P., Nerry, F., Ottlé, C., 2004. Land surface temperature retrieval techniques and applications: case of the AVHRR. In: Thermal Remote Sensing in Land Surface Processes. CRC Press, pp. 33–109.
- Kljun, N., Calanca, P., Rotach, M., Schmid, H., 2004. A simple parameterisation for flux footprint predictions. Bound.-Layer Meteorol. 112, 503–523. <https://doi.org/10.1007/s10653-004-0196-2>.
- Lacerda, L.N., Cohen, Y., Snider, J., Huryna, H., Liakos, V., Vellidis, G., 2021. Field scale assessment of the TsHARP technique for thermal sharpening of MODIS satellite images using VENUS and Sentinel-2-derived NDVI. Remote Sens 13, 1155. <https://doi.org/10.3390/rs13061155>.
- Leander, R., Buishand, T.A., 2007. Resampling of regional climate model output for the simulation of extreme river flows. J. Hydrol. 332, 487–496. <https://doi.org/10.1016/j.jhydrol.2006.08.006>.
- Li, K., Guan, K., Jiang, C., Wang, S., Peng, B., Cai, Y., 2021a. Evaluation of four new land surface temperature (LST) products in the US corn belt: ECOSTRESS, GOES-R, Landsat, and Sentinel-3. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 9931–9945. <https://doi.org/10.1109/JSTARS.2021.3114613>.
- Li, S., Wang, J., Li, D., Ran, Z., Yang, B., 2021b. Evaluation of Landsat 8-like land surface temperature by fusing Landsat 8 and MODIS land surface temperature product. Processes 9, 2262. <https://doi.org/10.3390/pr9122262>.
- Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013. Satellite-derived land surface temperature: current status and perspectives. Remote Sens. Environ. 131, 14–37. <https://doi.org/10.1016/j.rse.2012.12.008>.
- Liu, H., Wang, Q., 2018. Scaling effect of fused ASTER-MODIS land surface temperature in an urban environment. Sensors 18, 4058. <https://doi.org/10.3390/s18114058>.
- Liu, K., Su, H., Li, X., Wang, W., Yang, L., Liang, H., 2016. Quantifying spatial-temporal pattern of urban heat island in Beijing: an improved assessment using land surface temperature (LST) time series observations from LANDSAT, MODIS, and Chinese new satellite GaoFen-1. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 2028–2042. <https://doi.org/10.1109/JSTARS.2015.2513598>.
- Liu, L., Zhang, Y., 2011. Urban heat island analysis using the landsat TM data and ASTER data: a case study in Hong Kong. Remote Sens 3, 1535–1552. <https://doi.org/10.3390/rs3071535>.
- Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., Shi, C., 2019. Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. Remote Sens. Environ. 233, 111364. <https://doi.org/10.1016/j.rse.2019.111364>.
- Long, D., Singh, V.P., 2012a. A modified surface energy balance algorithm for land (M-SEBAL) based on a trapezoidal framework. Water Resour. Res. 48 <https://doi.org/10.1029/2011WR010607>.
- Long, D., Singh, V.P., 2012b. A two-source trapezoid model for evapotranspiration (TTME) from satellite imagery. Remote Sens. Environ. 121, 370–388. <https://doi.org/10.1016/j.rse.2012.02.015>.
- Long, D., Singh, V.P., 2013. Assessing the impact of end-member selection on the accuracy of satellite-based spatial variability models for actual evapotranspiration estimation. Water Resour. Res. 49, 2601–2618. <https://doi.org/10.1002/wrcr.20208>.
- Long, D., Yan, L., Bai, L., Zhang, C., Li, X., Lei, H., Yang, H., Tian, F., Zeng, C., Meng, X., 2020. Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach. Remote Sens. Environ. 246, 111863. <https://doi.org/10.1016/j.rse.2020.111863>.
- Lymburner, L., Tan, P., McIntyre, A., Thankappan, M., Sixsmith, J., 2015. Dynamic land cover dataset version 2.1 [Dataset]. Available at: Geoscience Australia, Canberra. Accessed 20/Mar/2023. <http://pid.geoscience.gov.au/dataset/ga/83868>.
- Ma, J., Shen, H., Wu, P., Wu, J., Gao, M., Meng, C., 2022. Generating gapless land surface temperature with a high spatio-temporal resolution by fusing multi-source satellite-observed and model-simulated data. Remote Sens. Environ. 278, 113083. <https://doi.org/10.1016/j.rse.2022.113083>.
- Ma, J., Zhang, W., Marinoni, A., Gao, L., Zhang, B., 2018. Performance assessment of ESTARFM with different similar-pixel identification schemes. J. Appl. Remote. Sens. 12, 025017. <https://doi.org/10.1111/j.1365-0420.2017.025017>.
- McColl, K.A., Vogelzang, J., Konings, A.G., Entekhabif, D., Piles, M., Stoffelen, A., 2014. Extended triple collocation: estimating errors and correlation coefficients with respect to an unknown target. Geophys. Res. Lett. 41, 6229–6236. <https://doi.org/10.1002/2014GL061322>.
- McVicar, T.R., Jupp, D.L., 1999. Estimating one-time-of-day meteorological data from standard daily data as inputs to thermal remote sensing based energy balance models. Agric. For. Meteorol. 96, 219–238. [https://doi.org/10.1016/S0168-1923\(99\)00052-0](https://doi.org/10.1016/S0168-1923(99)00052-0).
- Mohamed, A.A., Odindi, J., Mutanga, O., 2017. Land surface temperature and emissivity estimation for urban Heat Island assessment using medium-and low-resolution space-borne sensors: a review. Geocarto International 32, 455–470. <https://doi.org/10.1080/10106049.2016.1155657>.
- Montanaro, M., Gerace, A., Lunsford, A., Reuter, D., 2014. Stray light artifacts in imagery from the landsat 8 thermal infrared sensor. Remote Sens. 6, 10435–10456. <https://doi.org/10.3390/rs61110435>.
- 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>.
- Mudele, O., 2021. pylandtemp: A Python package for computing land surface temperature from Landsat satellite imagery. Available at: GitHub. Accessed 19/Mar/2023. <https://github.com/pylandtemp/pylandtemp>.
- Parastatidis, D., Mitraka, Z., Chrysoulakis, N., Abrams, M., 2017. Online global land surface temperature estimation from landsat. Remote Sens 9, 1208. <https://doi.org/10.3390/rs9121208>.
- Pu, R., Bonafoni, S., 2023. Thermal infrared remote sensing data downscaling investigations: An overview on current status and perspectives. Remote Sens. Appl. Soc. Environ. 100921 <https://doi.org/10.1016/j.rsase.2023.100921>.

- Quan, J., Zhan, W., Ma, T., Du, Y., Guo, Z., Qin, B., 2018. An integrated model for generating hourly landsat-like land surface temperatures over heterogeneous landscapes. *Remote Sens. Environ.* 206, 403–423. <https://doi.org/10.1016/j.rse.2017.12.003>.
- Reuter, D.C., Richardson, C.M., Pellerano, F.A., Irons, J.R., Allen, R.G., Anderson, M., Jhabvala, M.D., Lunsford, A.W., Montanaro, M., Smith, R.L., 2015. The thermal infrared sensor (TIRS) on landsat 8: design overview and pre-launch characterization. *Remote Sens* 7, 1135–1153. <https://doi.org/10.3390/rs70101135>.
- Roy, D.P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of landsat data. *Remote Sens. Environ.* 112, 3112–3130. <https://doi.org/10.1016/j.rse.2008.03.009>.
- Seemann, S.W., Borbas, E.E., Knuteson, R.O., Stephenson, G.R., Huang, H.-L., 2008. Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance measurements. *J. Appl. Meteorol. Climatol.* 47, 108–123. <https://doi.org/10.1175/2007JAMC1590.1>.
- Semmens, K.A., Anderson, M.C., Kustas, W.P., Gao, F., Alfieri, J.G., McKee, L., Prueger, J.H., Hain, C.R., Cammalleri, C., Yang, Y., 2016. Monitoring daily evapotranspiration over two California vineyards using landsat 8 in a multi-sensor data fusion approach. *Remote Sens. Environ.* 185, 155–170. <https://doi.org/10.1016/j.rse.2015.10.025>.
- Shen, H., Meng, X., Zhang, L., 2016. An integrated framework for the spatio-temporal-spectral fusion of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 54, 7135–7148. <https://doi.org/10.1109/TGRS.2016.2596290>.
- Shi, C., Wang, N., Zhang, Q., Liu, Z., Zhu, X., 2022. A comprehensive flexible spatiotemporal DAta fusion method (CFSDAF) for generating high spatiotemporal resolution land surface temperature in urban area. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 15, 9885–9899. <https://doi.org/10.1109/JSTARS.2022.3220897>.
- Van Niel, T., McVicar, T.R., Fang, H., Liang, S., 2003. Calculating environmental moisture for per-field discrimination of rice crops. *Int. J. Remote Sens.* 24, 885–890. <https://doi.org/10.1080/0143116021000009921>.
- Van Niel, T.G., McVicar, T.R., 2004. Determining temporal windows for crop discrimination with remote sensing: a case study in South-Eastern Australia. *Comput. Electron. Agric.* 45, 91–108. <https://doi.org/10.1016/j.compag.2004.06.003>.
- Vergopolan, N., Chaney, N.W., Beck, H.E., Pan, M., Sheffield, J., Chan, S., Wood, E.F., 2020. Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sens. Environ.* 242, 111740. <https://doi.org/10.1016/j.rse.2020.111740>.
- Vlassova, L., Perez-Cabello, F., Nieto, H., Martín, P., Riaño, D., De La Riva, J., 2014. Assessment of methods for land surface temperature retrieval from Landsat-5 TM images applicable to multiscale tree-grass ecosystem modeling. *Remote Sens* 6, 4345–4368. <https://doi.org/10.3390/rs6054345>.
- 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.2013.08.027>.
- Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sens. Environ.* 204, 31–42. <https://doi.org/10.1016/j.rse.2017.10.046>.
- Weng, Q., Fu, P., Gao, F., 2014. Generating daily land surface temperature at landsat resolution by fusing landsat and MODIS data. *Remote Sens. Environ.* 145, 55–67. <https://doi.org/10.1016/j.rse.2014.02.003>.
- Xia, H., Chen, Y., Li, Y., Quan, J., 2019. Combining kernel-driven and fusion-based methods to generate daily high-spatial-resolution land surface temperatures. *Remote Sens. Environ.* 224, 259–274. <https://doi.org/10.1016/j.rse.2019.02.006>.
- Yang, G., Weng, Q., Pu, R., Gao, F., Sun, C., Li, H., Zhao, C., 2016. Evaluation of ASTER-like daily land surface temperature by fusing ASTER and MODIS data during the HiWATER-MUSOEXE. *Remote Sens* 8, 75. <https://doi.org/10.3390/rs8010075>.
- Yamamoto, Y., Ishikawa, H., Oku, Y., Hu, Z., 2018. An algorithm for land surface temperature retrieval using three thermal infrared bands of Himawari-8. *J. Meteorol. Soc. Jpn Ser. II* 96B, 59–76. <https://doi.org/10.2151/jmsj.2018-005>.
- Yang, Y., Guan, H., Long, D., Liu, B., Qin, G., Qin, J., Batelaan, O., 2015. Estimation of surface soil moisture from thermal infrared remote sensing using an improved trapezoid method. *Remote Sens* 7, 8250–8270. <https://doi.org/10.3390/rs70708250>.
- Yu, Y., Renzullo, L.J., McVicar, T.R., Van Niel, T.G., Cai, D., Tian, S., 2024. Zenith Angle-based Calibration (ZAC) of Himawari-8 land surface temperature based on MODIS spatiotemporal characteristics. *In Preparation*.
- Yu, Y., Renzullo, L.J., Tian, S., 2021. Continental scale downscaling of AWRA-L analysed soil moisture using random forest regression. In: MODSIM2021, 24th International Congress on Modelling and Simulation, Sydney, Australia, 5–10 December. <https://doi.org/10.36334/modsim.2021.J10.yu>.
- Yuan, F., Bauer, M.E., 2007. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in landsat imagery. *Remote Sens. Environ.* 106, 375–386. <https://doi.org/10.1016/j.rse.2006.09.003>.
- Zhan, W., Huang, F., Quan, J., Zhu, X., Gao, L., Zhou, J., Ju, W., 2016. Disaggregation of remotely sensed land surface temperature: a new dynamic methodology. *J. Geophys. Res. Atmos.* 121, 10538–10554. <https://doi.org/10.1002/2016JD024891>.
- Zhang, L., Jiao, W., Zhang, H., Huang, C., Tong, Q., 2017. Studying drought phenomena in the continental United States in 2011 and 2012 using various drought indices. *Remote Sens. Environ.* 190, 96–106. <https://doi.org/10.1016/j.rse.2016.12.010>.
- Zhang, W., Li, A., Jin, H., Bian, J., Zhang, Z., Lei, G., Qin, Z., Huang, C., 2013. An enhanced spatial and temporal data fusion model for fusing landsat and MODIS surface reflectance to generate high temporal landsat-like data. *Remote Sens* 5, 5346–5368. <https://doi.org/10.3390/rs5105346>.
- Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* 114, 2610–2623. <https://doi.org/10.1016/j.rse.2010.05.032>.
- Zhu, X., Helmer, E.H., Gao, F., Liu, D., Chen, J., Lefsky, M.A., 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* 172, 165–177. <https://doi.org/10.1016/j.rse.2015.11.016>.
- Zhu, X., Song, X., Leng, P., Li, X., Gao, L., Guo, D., Cai, S., 2021. A framework for generating high spatiotemporal resolution land surface temperature in heterogeneous areas. *Remote Sens* 13, 3885. <https://doi.org/10.3390/rs13193885>.