

# *Supplement of the radiance-based method to validate satellite-derived land surface temperature products over heterogeneous land surfaces*

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Yu, W., Ma, M., Yang, H. ORCID: <https://orcid.org/0000-0001-9940-8273>, Tan, J. and Li, X. (2019) Supplement of the radiance-based method to validate satellite-derived land surface temperature products over heterogeneous land surfaces. *Remote Sensing of Environment*, 230. 111188. ISSN 0034-4257 doi: <https://doi.org/10.1016/j.rse.2019.05.007> Available at <https://centaur.reading.ac.uk/84316/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.rse.2019.05.007>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Supplement of the radiance-based method to validate satellite-derived land surface temperature products over heterogeneous land surfaces

Wenping Yu <sup>a, b</sup> \*, Mingguo Ma <sup>a, b</sup>, Hong Yang <sup>d</sup>, Junlei Tan <sup>c</sup>, and Xiaolu Li <sup>a, b</sup>

<sup>a</sup> Chongqing Engineering Research Center for Remote Sensing Big Data Application, School of Geographical Sciences, Southwest University, No. 2 Tiansheng Road, Beibei District, Chongqing 400715, China

<sup>b</sup> Research Base of Karst Eco-environments at Nanchuan in Chongqing, Ministry of Nature Resources, School of Geographical Sciences, Southwest University, Chongqing 400715, China

<sup>c</sup> Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, 320 Donggang West Road, Lanzhou 730000, China

<sup>d</sup> Department of Geography and Environmental Science, University of Reading, Whiteknights, Reading, RG6 6AB, UK

\* Correspondence: [ywpgis2005@swu.edu.cn](mailto:ywpgis2005@swu.edu.cn)

## Abstract

Land surface temperature (LST) retrieved from satellite remote sensing data has become a key parameter in research on global environmental change; therefore, the acquisition of accurate satellite-derived LST information is crucial for the diagnosis and analysis of global change. However, it is relatively difficult to obtain the true value of a pixel due to the scale mismatch between in situ measurements and satellite-based observations, especially for commonly heterogeneous and nonisothermal land areas, which greatly increases the difficulty in estimating pixel-representative LST values from in situ measurements for validation of satellite-based LST products. In this study, a supplemented radiance-based (SR-based) validation method was developed to evaluate the latest moderate resolution imaging spectroradiometer (MODIS) Collection 6 Level 2 daily LST/land surface emissivity (LSE) products over a heterogeneous and nonisothermal region of the Heihe Watershed Allied Telemetry Experimental Research (HiWATER) project, West China. In the SR-based framework, pixel-representative LST values are simulated by the MODTRAN model from the corresponding in situ measurements, such as LSE and atmospheric profile measurements, to evaluate the MODIS LST products. The validation results show that the MODIS daytime LST products from the Aqua satellite (MYD11\_L2) have a greater accuracy than those from the Terra satellite (MOD11\_L2). Analyses of the effect factors indicate a strong correlation between the errors in the MOD11\_L2 LST product and the corresponding difference in the MODIS brightness temperature between bands 31 and 32. Although the requirement of synchronous or quasisynchronous in situ measurements for the validated LST products may limit the applicability of the SR-based method, it is still an effective and simple method for validating satellite-derived LST products over mixed pixels. Our method is an indispensable supplement for the validation methods of satellite-derived LST products, and it can be applied in West China and other areas with heterogeneous land surfaces.

**Keywords:** heterogeneous land surface, validation, land surface temperature, MODIS, HiWATER

---

## 1. Introduction

The land surface temperature (LST) is a parameter of utmost importance in the global water and energy cycle (Anderson et al., 2012; Brunsell and Gillies, 2003; Li et al., 2013b; Rozenstein et al., 2014). Moreover, knowledge of the LST is indispensable for developing an understanding of the spatial variations in the surface energy balance, which is essential in various fields, such as drought assessment, vegetation monitoring, and climatic change and hydrological cycle research (Arnfield, 2003; Bastiaanssen et al., 2005; Huang et al., 2018; Kogan, 2001; Weng, 2009; Zhao et al., 2017). With the evolution of space science and technologies, satellite-based LST data have provided the only opportunity of measuring LSTs at the global scale with sufficiently spatiotemporal resolution. Compared to traditional point measurements, satellite-based LST data provide spatial average values. Retrieving LSTs from aerospace-based thermal infrared (TIR) data has attracted much attention since the 1970s (McMillin, 1975), and numerous satellite-based operational LST products have been developed based on the TIR data from a variety of satellite platforms (Prata, 2002; Smith et al., 1997; Trigo et al., 2008b; Wan and Dozier, 1996; Zhang et al., 2019). Among these LST products, the moderate resolution imaging spectroradiometer (MODIS) LST dataset is one of the essential LST datasets and presently widely utilized.

Without validation, no algorithms, simulations, or parameters retrieved from remote sensing data can be applied sufficiently accurately. Therefore, the independent validation process has become an indispensable step for the application of these satellite-derived LSTs, which can provide reliable information to potential users and feedback to LST algorithm developers. Until now, three technologies have been generally utilized to evaluate satellite-derived LST data (Li et al., 2013b). The first one is the temperature-based (T-based) method, which directly calculates the difference between the satellite-based LST and the synchronous ground-based LST (Coll et al., 2005; Martin et al., 2019; Pinker et al., 2009; Slater et al., 1996; Wan, 2014; Wan et al., 2002). The second one is the radiance-based (R-based) method, which compares the satellite-based LST to the adjusted LST. According to the ground-based atmospheric profiles and land surface emissivity (LSE) data corresponding to the satellite overpass, this method can simulate the top of the atmosphere (TOA) radiation, thereby matching the satellite-observed radiation based on a given radiation transfer model (RTM) (Coll et al., 2012a; Coll et al., 2012b; Coll et al., 2009; Duan et al., 2018; Wan, 2014; Wan and Li, 2008; Wan et al., 2002). The third one is the cross-validation method, which uses the LST derived from the algorithm verified with complete documentation to assess other satellite-derived LST values. The cross-validation method is a replacement for the T-based and R-based methods when atmospheric profiles or ground-based LSTs are unavailable (Trigo et al., 2008a). Due to the difficulty of matching ground-based LST measurements with satellite-based LST pixel values, the T-based method is limited to homogeneous and flat surfaces, including inland water, snow, sand, ice, grasslands, and agriculture. Compared with the T-based method, the R-based method has an obvious advantage for validation when ground-based LST measurements are unfeasible because it provides the opportunity to extend the validations over heterogeneous surfaces (Li et al., 2013b; Yu et al., 2017).

A series of studies have validated satellite-derived LST products over largely homogenous land (Coll et al., 2009; Li et al., 2013b; Wan and Li, 2008; Wan et al., 2004; Yu et al., 2017). Up to now, to the best of our knowledge, few studies have evaluated satellite-derived LST products over heterogeneous surfaces, due to the difficulty of measuring representative ground-based LSTs at the satellite pixel scale. However, heterogeneous land surface is the usual case in Earth observation data. The systematic site-to-network (SSN) is an acceptable system strategy, supported by multisampling strategies and scale transfer technologies, to obtain the “true” LSTs over mixed pixels on the basis of multiscale, multiplatform and multisource measurements (Wang et al., 2016). However, few referable LST images are available at satellite overpasses, and they are limited by the cost and repeated observations of the remote sensors with high spatial resolutions. Moreover, the validity of the SSN validation method depends on in situ sampling and scale transfer issues, involving upscaling and downscaling methods, a fundamental question for quantitative remote sensing research. All of these issues make it complex and difficult to implement the SSN method.

## 2. Methodology

Due to the spatial resolution of a TIR sensor aboard a satellite is always coarser than hundreds of meters or even kilometers, the pixels in TIR images are usually composites of several scene components. Based on the assumption that the ensemble radiance of a given mixed pixel is linearly influenced by the radiance of each element emitted in the pixel, the mixed pixel radiance could be weighted by the endmembers according to their projected area fractions in the pixel (Ermiida et al., 2014; Pinheiro et al., 2006). Therefore, based on the theory of Pinheiro et al. (2006), for a given area composed of  $N$  scene components (endmembers), the general formulation of this relationship is as follows:

$$L_{area} = \sum_{k=1}^N (L_k f_k) \quad (1)$$

where  $L_{area}$  is the ensemble radiance of a given area;  $L_k$  is the radiance emitted by endmember  $k$ ; and  $f_k$  is the corresponding component fraction. Assuming endmembers are homogeneous elements, according to the Stefan-Boltzmann law, the LST of the simulated pixel can be derived as Eq. (2) from Eq. (1):

$$T_{area} = \sqrt[4]{\frac{\sum_{k=1}^N \varepsilon_k f_k T_k^4}{\varepsilon_{area}}} \quad (2)$$

where  $\varepsilon_k$  is the emissivity of endmember  $k$ , which can be measured concurrently;  $T_k$  is the LST of endmember  $k$  over the heterogeneous area; and  $\varepsilon_{area}$  is the effective emissivity of the given area. According to Equation (2), the LST over a heterogeneous area depends on the component emissivities, effective emissivity, cover fractions, and component LSTs. In Equation (2), the key variable  $T_k$  is difficult to measure concurrently and representatively because of its strong spatiotemporal heterogeneities. Thus, the T-based validation method based on ground-based LST measurements cannot be applied to validate LSTs over heterogeneous pixels.

The R-based method proposed by Wan and Li (2008) supplies a validation strategy suitable for evaluating satellite-derived LSTs globally on the independence of ground-based LST measurements. The ‘‘true’’ LST of a pixel is simulated on the reference of the TOA radiance of the validated sensor by a selected RTM code based on the inputs of spectral LSE data for given land surfaces, atmospheric water vapor (AWV) and temperature profiles concurrent with satellite. Therefore, the R-based method represents an advanced technique to validate satellite-derived LSTs. An extended radiance-based method (ER-based) is proposed in this study to validate satellite-derived LST data over heterogeneous pixels with the R-based method as a basis. Figure 1 is the flow diagram of the ER-based method. MODTRAN 5.2 (Berk et al. 2005) was employed in the ER-based validation. Because of difficulty of describing the directional natural emissivity, the emissivity has been assumed to be isotropic in most previous studies (Ermida et al. 2014; Pinheiro et al. 2006). Therefore, the effective emissivity of the given area can be calculated as follows (Pinheiro et al. 2006):

$$\varepsilon_{area} = \sum_{k=1}^N (\varepsilon_k * f_k) \quad (3)$$

where  $f_k$  is calculated from the proportions of the surface covered by different components. The fractional vegetation cover (FVC) is also helpful to obtain  $\varepsilon_k$  in an incompletely vegetated endmember  $k$ . In this study, to extend Equation (3), the temperature-independent effective spectral emissivity of a given area can be defined under the condition that other surface attributes do not vary with the LST as follows (Ermida et al., 2014; Pinheiro et al., 2006; Wan and Dozier, 1996)::

$$\varepsilon_{\lambda,area} = \sum_{k=1}^N (\varepsilon_{\lambda,k} * f_k) \quad (4)$$

where  $\varepsilon_{\lambda,area}$  is the area effective spectral emissivity at wavelength  $\lambda$  and  $\varepsilon_{\lambda,k}$  is the emissivity of endmember  $k$  at wavelength  $\lambda$ . The effective spectral emissivity of a pixel over a heterogeneous surface is considered as a weighted mean of the component emissivities on the basis of concurrent component temperature-independent spectral emissivity measurements, FVC measurements, and accurate land cover type (LCT) images with high spatial resolutions (Figure 1). In the ER-based method, the effective spectral emissivity of a pixel is used as the spectral emissivity input for the RTM (MODTRAN 5.2). To simulate the TOA TIR band radiance obtained by the sensors at satellite overpasses, the satellite-based LST and ground-based profiles of atmospheric parameters extracted from the concurrent *in situ* radiosonde measurements are also treated as initial input parameters. Based on a comparison of the radiances obtained from the satellite-based observations and RTM-based simulations, the adjusted LST from the initial LST is inputted to iteratively calculate the TOA radiation that matches the radiation received by the TIR sensor. Then, the adjusted LST is considered as the ‘‘true’’ LST, and the accuracy of the satellite-derived LST is estimated by the difference between itself and the adjusted LST.

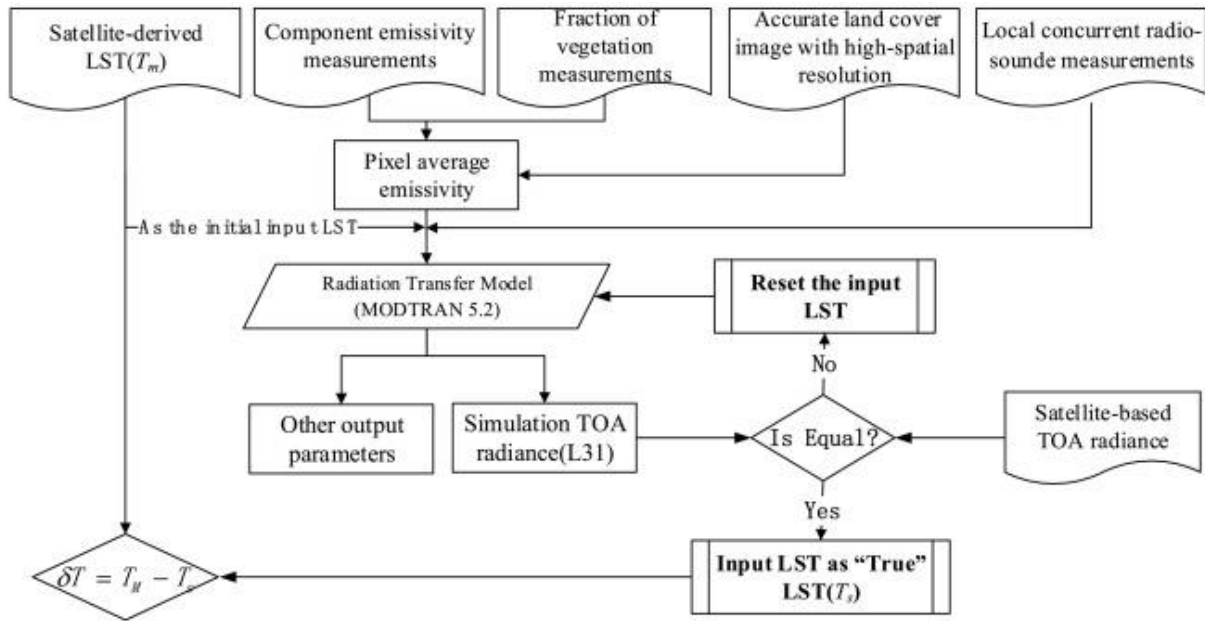


Figure 1. Flowchart of the ER-based validation in this study.

### 3. Data Description and Preparation

This research was conducted in Heihe River Basin (HRB) (37.7°-42.7°N, 97.1°-102.0°E) in the northwestern arid zone of China. The river basin area is approximately 1,432,000 km<sup>2</sup> (Li et al. 2013a). Landscape varies from glaciers to desert from upstream to downstream, including various intermediate zones (Figure 2). The HRB is also a typical exemplification of an inland river basin in the northwestern arid area of China, suffering from water-stress, climate change and human intervention. Therefore, the HRB was selected to carry out for land surface or hydrological experiments (Cheng 2009), for example, the Watershed Allied Telemetry Experimental Research (i.e., WATER) (Li et al. 2009) and Heihe Watershed Allied Telemetry Experimental Research (HiWATER) (Huang et al. 2016; Li et al. 2013a; Li et al. 2017; Yu et al. 2011). To monitor the long-term change in the HRB, three key experimental areas (KEA) plans were implemented in the upstream (cold zone), midstream (artificial oasis zone) and downstream (natural oasis zone). In this study, the validation of satellite-derived LSTs was mainly implemented in the second KEA. A zone with dimensions of approximately 5 km × 5 km was selected as the foci experimental area (FEA) (the black rectangular area in Figure 2) in the middle KEA where 17 automatic meteorological stations were installed. To capture the strong land surface heterogeneities of the corresponding remote sensing products, airborne remote sensing and ground-based remote sensing observations as well as some auxiliary ground-based measurements were obtained at the same time as the satellite overpasses. The remote sensing data and ground-based measurements involved in this study are detailed below.

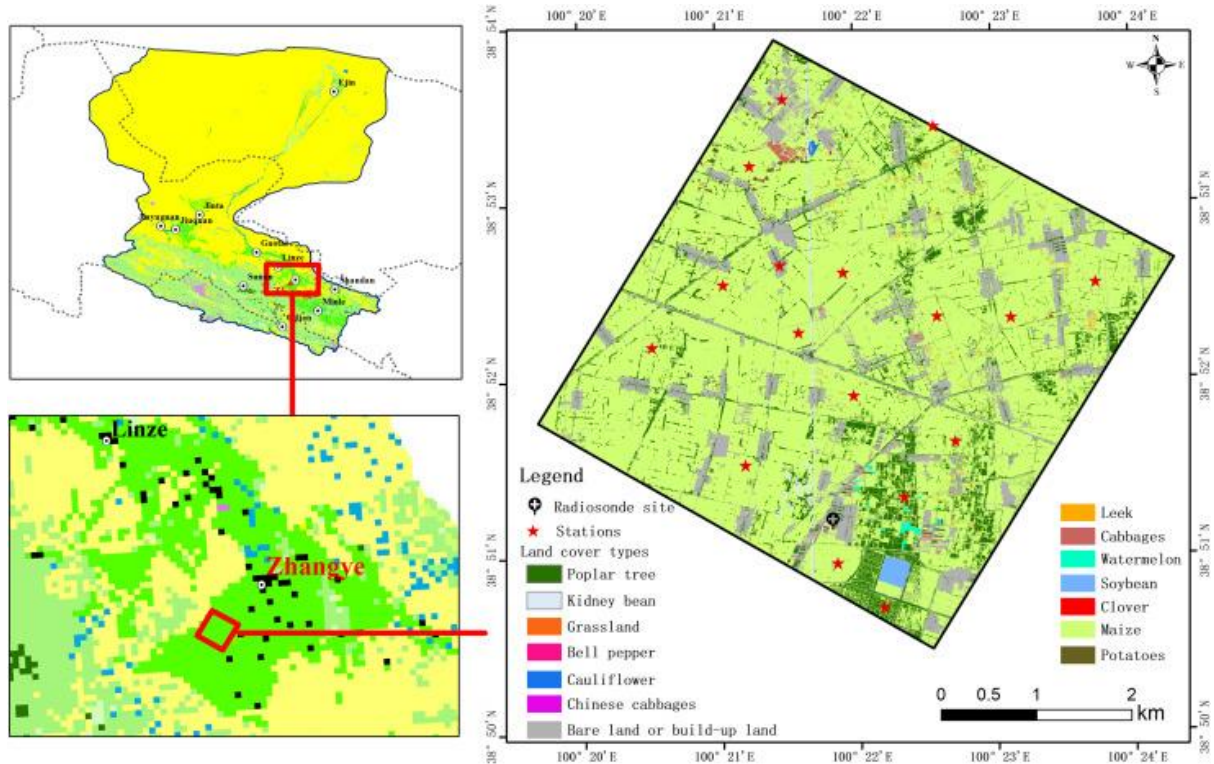


Figure 2. Location of the foci experimental area (FEA) in the Heihe River Basin (HRB) in Northwest China.

### 3.1. Remote sensing data

The Collection 6 LST data retrieved from MODIS onboard the Terra (morning) and Aqua (afternoon) satellites have been available since 2000 and 2002, respectively (Salomonson et al., 1989). Depending on the TIR data of MODIS, a series of MODIS LSTs released in a MODIS product list (the website is [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table)), and the refined collection products are constantly updated (Li et al., 2013b; Wan, 2014; Wan and Li, 1997; Wan et al., 2002). MOD11\_L2 and MYD11\_L2, which are the basic LST products and have a 1-km spatial resolution, were retrieved by the generalized split-window (GSW) method proposed by Wan and Dozier (1996). The Collection 6 LST products are the latest version of the MODIS LSTs. Compared with Collection 5, the refinements of Collection 6 can be summarized in three points (Wan, 2014): the use of two different parameter settings for hot and warm bare soil to refine the GSW method for pixels; the addition of an emissivity adjustment of the same magnitude but opposite direction to the bare soil emissivity in bands 31 and 32; and the incorporation of new sets of GSW coefficients in the day/night algorithm (MOD11B1 and MYD11B1) with a 5-km spatial resolution (Wan and Li, 1997). Since the latest MODIS Collection 6 LST products were released, little research has focused on their evaluation. Because MOD11\_L2 and MYD11\_L2 are the basis of MOD11A1 and MYD11A1 (gridded LST products from MOD11\_L2 and MYD11\_L2) and have a higher spatial resolution than MOD11B1 and MYD11B1, the daytime MOD11\_L2 and MYD11\_L2 were selected as the satellite-derived LST product example to implement the SR-based validation process.

According to Figure 1, the “true” LST of a heterogeneous pixel is simulated based on the received radiance of the remote sensors. For the validation of the MODIS LST products, the radiance of MODIS band 31 with a wavelength from 10.78  $\mu\text{m}$  to 11.28  $\mu\text{m}$  was selected to calculate the “true” pixel LST of the MODIS depending on the fact that the radiation in MODIS band 31 spectrum is least affected by the variations in AWP and temperature profiles (Coll et al. 2009; Wan and Li 2008). Additionally, the uncertainty in the LSE in band 31 has been tested to exhibit small variations of the major landscapes (Snyder et al. 1998). To obtain the geographic coordinate information of the MODIS radiance data (MOD/MYD021KM), the corresponding location data (MOD/MYD03) were obtained from NASA (<https://ladsweb.modaps.eosdis.nasa.gov/search/>).

As shown in Equation (4) and the flowchart of the ER-based method (Figure 1), a LCT image with a high spatial resolution over the FEA region is essential for the estimation of the effective pixel emissivity. The airborne remote sensing experiment carried out in the HiWATER program provided the opportunity to obtain data with high spatial resolution at the time when the satellite-based remote sensor passed over. The spectral range of these airborne sensors extended from the visible and near-infrared (VNIR) bands to microwave bands. The LCT image in the FEA used in this study involves clipped LCT strips from LCT products in the middle observation KEA of

the HRB. The LCT products were retrieved by a hierarchical classification structure integrated with pixel-based classification and object-based classification from the CASI-1500 VNIR sensor (with a spectral range from 380 to 1050 nm, 48 bands, and a spatial resolution of 0.5 m) carried by a Harbin Y-12 aircraft. Data were collected on 29<sup>th</sup> June 2012. The LCT product has a spatial resolution of 1 m (<http://card.westgis.ac.cn/hiwater/vegtypeair>). The classification results of these products exhibited high precision, with a total accuracy of 84.61 % and a kappa coefficient of 0.8262. All vegetation types in the FEA LCT map were investigated in the field experiment, and the relevant data were recorded with the help of high-precision handheld GPS devices and digital cameras. Thus, the LCT map used here can provide an accurate representation of the proportions of surface covered by different components and can be used to obtain the effective pixel emissivity values.

### 3.2. Ground-based measurements

To investigate the upscaling issue of ecohydrological processes on complex underlying surfaces, a multiscale dataset was obtained from the HiWATER program in the middle KEA. All observations were obtained by previously calibrated instruments, and data quality control procedures were implemented throughout the original data measurement, preprocessing and dataset generation processes. The ground-based measurements used in this study were all obtained from the multiscale dataset freely available from the website of the Cold and Arid Regions Science Data Center (<http://card.westgis.ac.cn/>).

The atmospheric profiles were observed by Vaisala or Changfeng system radiosondes concurrent with the Terra or Aqua overpasses under clear-sky conditions, and the radiosonde launch sites in the FEA are shown in Figure 2. Table 1 gives the exact time that the balloons were launched in each synchronized field experiment. According to the weather records of the radiosonde observations, the data obtained on 7<sup>th</sup> July under light cloud conditions were excluded and other atmospheric profiles (Table 1) were inputted to the MODTRAN 5.2 code for the ER-based validation process. The atmospheric profiles were extended up to an altitude of 100 km with the midlatitude summer standard profile. Depending on the satellite viewing angle of each case, the path radiance and the atmospheric transmittance in the selected spectral range were simulated, and the sky flux was integrated over the hemisphere from the downwelling radiances from several angles. The observed spectral range was from 2 to 16  $\mu\text{m}$  with a spectral accuracy of  $1\text{ cm}^{-1}$  across the spectral range.

Table 1. Information on the launch of the radiosondes balloons.

Date	Time		Final altitude (m)	Observation system
	Start time	Stop time		
2012-06-29	11:40:10	12:57:09	16754	Changfeng system
2012-06-29	14:26:12	15:20:48	7008	Vaisala system
2012-06-30	13:54:23	16:38:49	37960	Vaisala system
2012-07-03	11:57:01	13:55:15	41154	Vaisala system
2012-07-04	11:24:04	13:03:28	32288	Vaisala system
2012-07-04	14:09:09	16:02:57	32384	Vaisala system
2012-07-05	11:06:33	13:09:09	38166	Vaisala system
2012-07-07	13:47:39	15:42:53	30220	Vaisala system
2012-07-08	13:49:18	15:27:02	33987	Vaisala system
2012-07-10	10:29:29	12:22:25	33491	Vaisala system
2012-07-10	14:45:03	16:28:47	32565	Vaisala system
2012-08-02	09:28:47	11:50:02	39354.1	Changfeng system
2012-08-02	13:55:09	16:12:45	14535.3	Changfeng system

The dataset used in this study includes the emissivity spectrum of typical ground objects in the FEA of the HRB (Li et al. 2017). The emissivity was measured by the Model 102 Portable Field Spectrometer (Hook and Kahle 1996; Hoover and Kahle 1987), a complete person-portable 102 model Fourier transform infrared (FT-IR) spectrometer (simply 102F). For the field measurements, the entire 102F instrument was completely sealed. To obtain the spectral emissivity data, cold and warm blackbody radiances were first measured to calibrate the instrument and eliminate the device noise. Then, the sample radiation  $L_s(\lambda)$  was measured and calibrated by the



cold and warm blackbody radiances. The downwelling calibrated radiation  $L_{dwr}(\lambda)$  of the samples was also obtained depending on the radiances from the diffuse reflectors. Thus, for each emissivity measurement, four raw data files were required: two blackbody radiances, one downwelling radiance and one sample itself. Finally, the absolute spectral emissivity of the target  $\varepsilon_{target}(\lambda)$  was computed as following:

$$\varepsilon_{target}(\lambda) = \frac{L_s(\lambda) - L_{dwr}(\lambda)}{B(\lambda, T) - L_{dwr}(\lambda)} \quad (5)$$

where  $B(\lambda, T)$  is the radiation based on the Planck function at the same temperature as the sample. To retrieve the emissivity and temperature from Equation (5), the linear emissivity constraint temperature and emissivity separation (LECTES) approach (Wang et al. 2011) was employed in this study. The emissivity spectrum was measured from 25<sup>th</sup> May to 18<sup>th</sup> July in 2012 concurrent with the aerial and satellite remote sensor overpasses.

Figure 2 shows the characteristics of the LCT in the FEA from May to September 2012. The change in the FVC in the FEA occurred mainly in the cropland areas, including vegetables and maize. Thus, the FVC dataset (doi:10.3972/hiwater.154.2013.db) includes primarily the FVC data generated from cropland. The FVC observations for vegetation growth cycles are from 9<sup>th</sup> May to 15<sup>th</sup> September in 2012 and are in five-day periods for each observation before 31<sup>st</sup> July and in ten-day periods for each observation after 31<sup>st</sup> July. In the field, digital photography measurements were implemented to measure the FVC, and a fixed-focus camera was mounted on the top of the platform with an adjustable length to conveniently observe different types of vegetation. At each crop observation point, 9 images along with two perpendicular cross-sections of the cropland were obtained, and the average value of the 9 FVC images was recognized as the “true FVC” of the given cropland.

The default aerosol models and meteorological ranges were usually used in the R-based method (Coll et al., 2009; Duan et al., 2018; Wan and Li, 2008), with the consequence of ignored effect of dust aerosols on the final validation results. Therefore, in the ER-based validation, the dataset of sun photometer observations in the FEA (the dataset doi:10.3972/hiwater.022.2013.db) were used to provide the atmospheric aerosol content and the column water vapor content for simulating the *in situ* LSTs. The datasets were measured by automatic sun tracking photometers (CE-318; NE type, produced by France Cimel) installed very close to the radiosonde site (Figure 1), and these instruments provided continuous data with an acquisition frequency of one minute from 1<sup>st</sup> June to 20<sup>th</sup> September in 2012. The dataset includes the raw observation data, retrieval data for the aerosol optical depth (AOD) in 9 bands and the AWV content retrieved from the band data with a center of 936 nm.

## 4. Results and Discussion

### 4.1. The validation results of the MOD/MYD11\_L2 LST products

In the ER-based validation, the “true” LSTs were simulated according to the TOA radiance in band 31 using the MODTRAN 5.2 radiative transfer model code depending on the local atmospheric profile measurements and the effective spectral emissivity of the MODIS pixels in the FEA. The difference between the MODIS LSTs (MOD11\_L2/MYD11\_L2, as  $T_{MOD}/T_{MYD}$ ) and the simulated “true” LST ( $T_{ER-based}$ ) was calculated to evaluate the accuracy of the MODIS Level 2 LST products. The validation results of MOD/MYD11\_L2 are shown in Table 2. To estimate the similarity between the pixel LSTs and the simulated *in situ* LSTs, a scatterplot containing the MOD/MYD11\_L2 pixels in the FEA and the simulated *in situ* LST  $T_{ER-based}$  is shown in Figure 3.

Table 2 indicates that the average errors of each image (the granule IDs and pixel numbers in the FEA are shown in the 1<sup>st</sup> and 2<sup>nd</sup> columns in Table 2) based on the difference ( $\overline{\delta T}$ ) between the pixel LST ( $T_{MOD}$ ) in MOD11\_L2 and the simulated *in situ* LST  $T_{ER-based}$  range from -1.79 K to 0.73 K, and the root-mean-square errors (RMSEs) range from 0.59 K to 1.81 K. Furthermore, the average errors of each image (the granule IDs and pixel numbers in the FEA are shown in the 6<sup>th</sup> and 7<sup>th</sup> columns in Table 2) based on the difference ( $\overline{\delta T}$ ) between the pixel LST ( $T_{MYD}$ ) in MYD11\_L2 and the simulated *in situ* LST ( $T_{ER-based}$ ) range from -0.5 K to 0.26 K, and the RMSEs range from 0.22 K to 0.57 K. The total validation results in daytime are  $-0.84 \pm 1.22$  K (daytime LST error mean  $\pm$  standard deviation) for the Terra MODIS Level 2 product and  $-0.11 \pm 0.43$  K for the Aqua MODIS Level 2 product. Additionally, from Figure 3(b), the LSTs ( $T_{MYD}$ ) of the MYD\_L2 pixels agree well with the *in situ* LSTs ( $T_{ER-based}$ ), and the linear fitting (the blue line) is very close to the  $y=x$  line (the red line). However, the overall LSTs ( $T_{MOD}$ ) of the MOD11\_L2 pixels are obviously underestimated relative to the corresponding *in situ* LSTs ( $T_{ER-based}$ ). Therefore, a noticeable accuracy difference exists between the new Collection 6 MOD11\_L2 and MYD11\_L2 products, and the LSTs from MYD11\_L2 have a greater accuracy in mixed pixels than those from MOD11\_L2.

According to the ER-based validation results mentioned above, the accuracy of the Collection 6 MYD11\_L2 products over mixed pixels is higher than that of the Collection 5 LST products validated by the R-based validation process in the homogeneous Hainich forest in Germany, with a mean bias of -3.0 K to 0.9 K and a RMSE of 0.2

K to 1.1 K(Wan 2014). While the MOD11\_L2 LSTs have a lower accuracy than that of the Collection 5 LST products over the homogeneous Hainich forest in Germany. Compared to the validation results for the MODIS Collection 6 Level 2 products (Wan 2014), the Aqua Level 2 daytime LST data over a mixed pixel in this study area has similar accuracy level as the Aqua Level 2 LST data (mean error=0.37 K) over a homogeneous bare soil site in Tamanrasset, Algeria, validated by the R-based process (Wan, 2014). In contrast, the accuracy of the Terra Level 2 daytime LST data over the mixed area is lower than that of the Terra Level 2 LST data for the bare soil site (mean error=0.51 K). Therefore, it reasonable to cautiously apply the Terra Level 2 daytime MODIS LSTs (MOD11\_L2) over mixed pixels directly.

Table 2. ER-based validation results for the MODIS LST products aboard the Terra and Aqua platforms.

MOD					MYD				
Granule ID	Pixel no.	$\tau_{31}$	$\overline{\delta T}(T_{MOD}-T_{ER-based})$	RMSE	Granule ID	Pixel no.	$\tau_{31}$	$\overline{\delta T}(T_{MYD}-T_{ER-based})$	RMSE
A2012181.0430	14	0.797	-0.98	0.99	A2012181.0610	13	0.831	-0.50	0.57
A2012185.0405	14	0.933	-0.25	0.59	A2012182.0650	12	0.911	0.02	0.22
A2012186.0445	13	0.920	0.73	0.84	A2012186.0625	14	0.917	0.19	0.55
A2012187.0350	14	0.816	-1.44	1.44	A2012190.0605	14	0.915	-0.45	0.49
A2012192.0410	14	0.922	-1.17	1.18	A2012192.0550	14	0.935	0.26	0.32
A2012215.0415	14	0.791	-1.79	1.81	A2012215.0555	14	0.786	-0.17	0.24

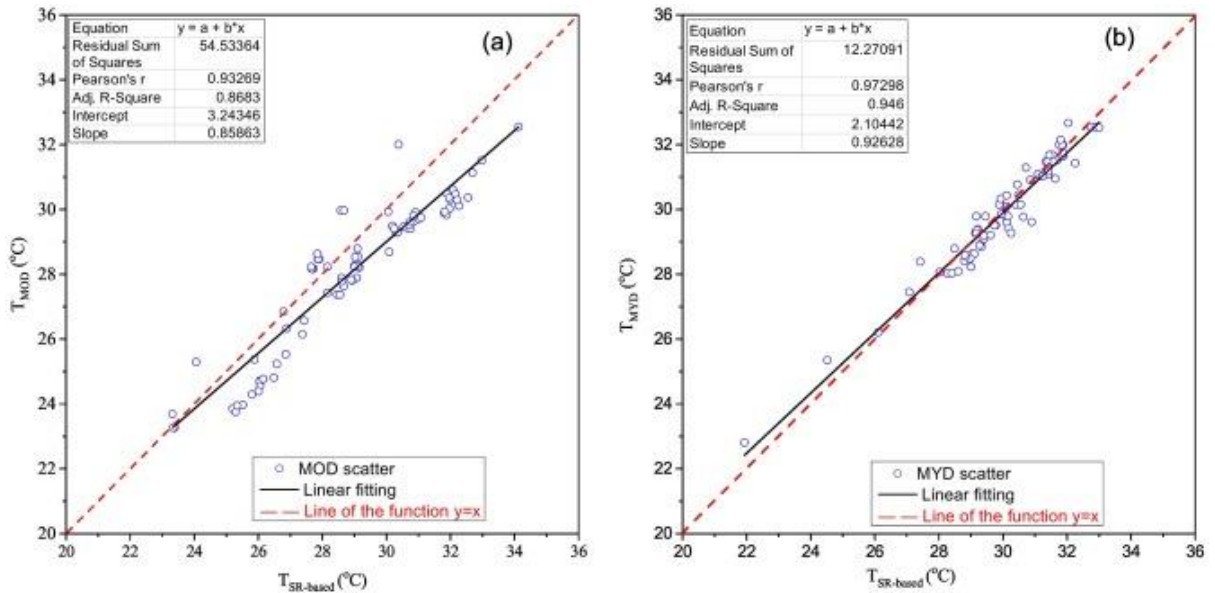


Figure 3. Scatterplots of the MODIS Level 2 LST and simulated *in situ* LST data.  $T_{MOD}$  in (a) is the Terra MODIS Level 2 daytime LSTs, and  $T_{MYD}$  in (b) is the Aqua MODIS Level 2 daytime LSTs.  $T_{ER-based}$  in (a) and (b) is the *in situ* LST corresponding to the MODIS LST pixel simulated by MODTRAN in the ER-based validation method.

#### 4.2. The possible factors responsible for the errors in MOD11\_L2

According to the validation results, the LST data from MOD11\_L2 have a lower accuracy than those from MYD11\_L2. Therefore, it is important to explore and identify the factors causing the LST accuracy difference between the MOD11\_L2 and MYD11\_L2 products. Wan (2014) developed the following refined split-window LST algorithm :

$$LST = b_0 + \left( b_1 + b_2 \frac{1 - \varepsilon}{\varepsilon} + b_3 \frac{\Delta \varepsilon}{\varepsilon^2} \right) \frac{T_i + T_j}{2} + \left( b_4 + b_5 \frac{1 - \varepsilon}{\varepsilon} + b_6 \frac{\Delta \varepsilon}{\varepsilon^2} \right) \frac{T_i - T_j}{2} + b_7 (T_i - T_j)^2 \quad (6)$$

where  $b_k$  ( $k=0\sim6$ ) represents the coefficients determined by a given viewing zenith angle (VZA) and for given subranges of  $\varepsilon$ , AWW, and the air temperature ( $T_a$ ) at the surface;  $T_i$  and  $T_j$  are the brightness temperatures in

bands  $i$  and  $j$ , respectively ( $i=31$  and  $j=32$ );  $\varepsilon$  is the mean emissivity of bands 31 ( $\varepsilon_{31}$ ) and 32 ( $\varepsilon_{32}$ ); and  $\Delta\varepsilon = \varepsilon_{31} - \varepsilon_{32}$ . Based on the equation, the LST depends mainly on  $b_k$ ,  $(T_i - T_j)$ ,  $(T_i + T_j)$ ,  $\varepsilon$ , and  $\Delta\varepsilon$ . The unknown coefficients are determined by VZA,  $\varepsilon$ , WV, and  $T_a$ . Therefore, the error analysis from the main terms in Equation (6) is an excellent place to start.

#### 4.2.1. The effect of the term $(T_{31} - T_{32})$ on the accuracy of MOD11\_L2

To explore the effect of the term  $(T_{31} - T_{32})$  on the accuracy of the Terra daytime MOD11\_L2 dataset, the relationship between the errors of all LSTs in the MOD11\_L2 pixels ( $T_{\text{MOD}} - T_{\text{ER-based}}$ ) and those in the MYD11\_L2 pixels ( $T_{\text{MYD}} - T_{\text{ER-based}}$ ), and their corresponding  $(T_{31} - T_{32})$  values from the MODIS radiance data (MOD021KM and MYD021KM) are shown in Figure 4. The errors of MOD11\_L2 have a significant relation with their corresponding  $(T_{31} - T_{32})$  values (Adjusted  $R^2 = 0.44$ , Pearson  $r = 0.664$  and  $p = 0.000 < 0.01$ ), but there is no significant relation between the errors of MYD11\_L2 and their corresponding  $(T_{31} - T_{32})$  values (Adjusted  $R^2 = 0.08$ , Pearson  $r = 0.299$  and  $p = 0.07 > 0.05$ ). The different effect of  $(T_{31} - T_{32})$  on MOD11\_L2 and MYD11\_L2 suggests that the term  $(T_{31} - T_{32})$  contributes to the accuracy difference between the daytime MOD11\_L2 and MYD11\_L2 products over heterogeneous land surfaces.

Furthermore, the distribution of the plots tend to converge on two linear trends. Therefore, the plots were split into two groups according to their distribution trends. Data analysis indicated that the two trends depend closely on the range of the atmospheric transmittance in band 31 ( $\tau_{31}$ ); and the  $\tau_{31}$  threshold between the plots with the two distribution trends is 0.8. Figure 5 shows the difference between the errors of MOD11\_L2 and their corresponding  $(T_{31} - T_{32})$  values for different ranges of  $\tau_{31}$ .  $\tau_{31}$  ranges from 0.75 to 0.80 in Figure 5(a), and from 0.8 to 0.94 in Figure 5(b). Figure 5 clearly shows that there is a significant linear relationship between the errors in MOD11\_L2 and the  $(T_{31} - T_{32})$  values (Adjusted  $R^2 > 0.7$ , Pearson  $r > 0.8$ ,  $p = 0.000 < 0.01$ ). Additionally, comparison between Figure 5(b) ( $R^2 = 0.94$ ) and Figure 5(a) (Adjusted  $R^2 = 0.72$  and Pearson  $r = 0.847$ ) indicates that the linear relationship increases with increasing  $\tau_{31}$ , and the accuracy of the MOD11\_L2 LST decreases by 2.42 K with a 1 K increase in the term  $(T_{31} - T_{32})$ .

Because the field experiments selected in this study were synchronously conducted with the airborne remote sensing experiments and MODIS sensor loaded on Terra or Aqua, the experiment data were all collected in fine weather conditions and the range of  $\tau_{31}$  in the SR-based validation was limited from 0.75 to 0.94. Therefore, in this study, we investigate the relationship between the errors in MOD11\_L2 and the term  $(T_{31} - T_{32})$  with respect to  $\tau_{31}$  values ranging from only 0.75 to 0.94. However, the results suggest that MOD/MYD11\_L2 should be more thoroughly validated before widespread application of the MODIS Collection 6 LST products with a 1-km spatial resolution, which are mostly based on the MOD/MYD11\_L2 products. Moreover, it is necessary to discuss the possible factors contributing to the greater errors of the MOD11\_L2 in this validation case.

Based on the analysis above, the error of MOD11\_L2 is closely related to the corresponding  $(T_{31} - T_{32})$  term and their linear relationship is restricted by the range of  $\tau_{31}$ ; therefore, the greater errors of MOD11\_L2 are likely associated with the inaccurate simulation of atmospheric conditions at the Terra passing time, which is used to determine the atmospheric transmittance and the optimal coefficients in the algorithm, especially for the coefficients of the  $(T_{31} - T_{32})$  term. The AWV,  $\text{CO}_2$ , and aerosol are the major factors in the atmosphere influencing the absorbing, scattering and emitting process in the TIR bands 31 and 32. Since the  $\text{CO}_2$  of atmosphere is almost constant, AWV and aerosol are key factors that may contribute to the greater errors of MOD11\_L2 based on the atmosphere effect of the GSW algorithm, which will be explained in the following two parts.

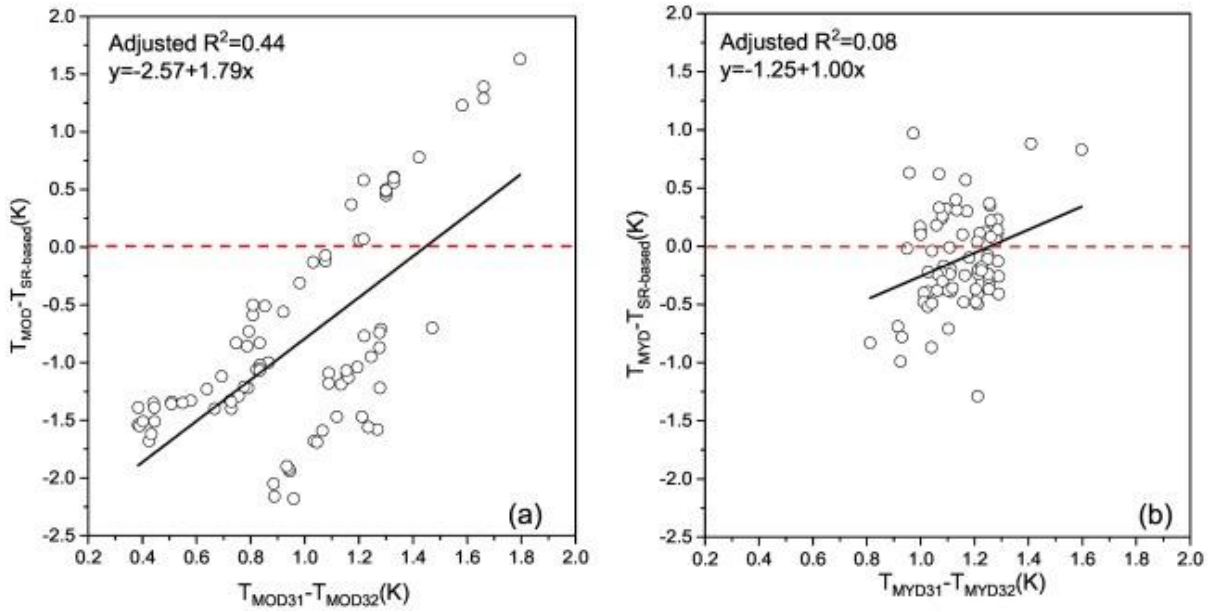


Figure 4. Relationship between the MODIS LST errors and the  $(T_{31} - T_{32})$  values; (a) the relationship between the LST errors in MOD11\_L2 and their corresponding  $(T_{31} - T_{32})$  values, and (b) the relationship between LST errors in MYD11\_L2 and their corresponding  $(T_{31} - T_{32})$  values.

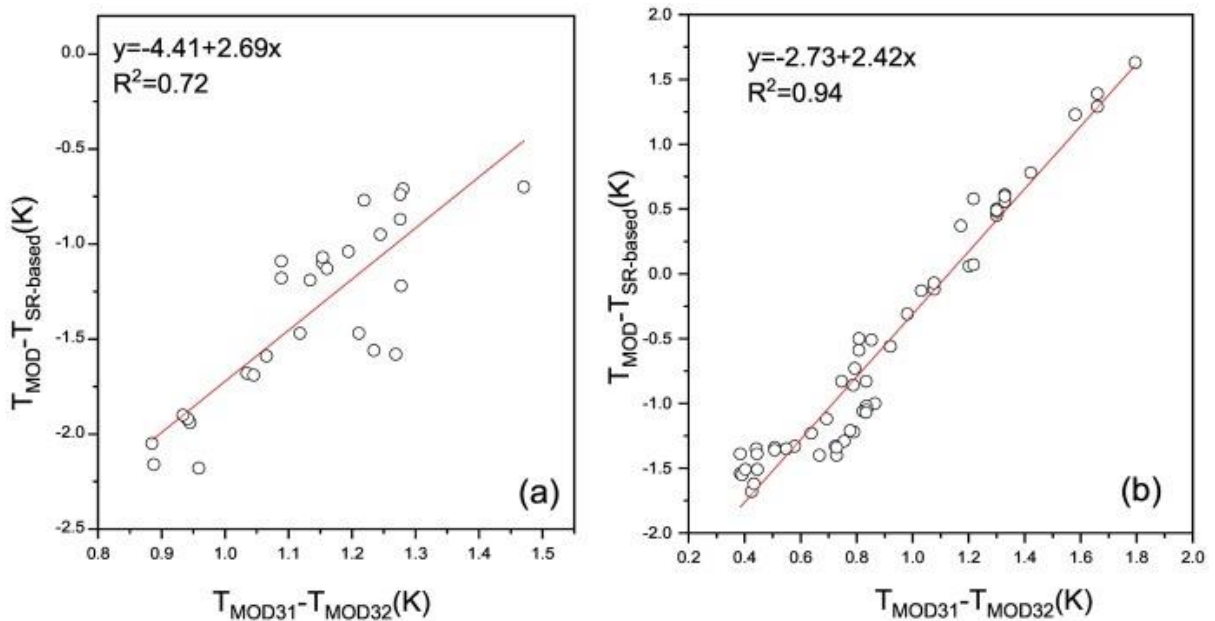


Figure 5. Relationship between the errors in the MOD11\_L2 product and their corresponding  $(T_{31} - T_{32})$  values for different ranges of  $\tau_{31}$ . (a),  $\tau_{31}$  ranges from 0.75 to 0.80; and (b),  $\tau_{31}$  is larger than 0.80.

#### 4.2.2. Impact of the AWV

To analyze the impact of the AWV accuracy on the error difference between MOD11\_L2 and MYD11\_L2, the corresponding MODIS Collection 6 daily atmospheric profile products (MOD/MYD07\_L2), with a spatial resolution of 5 km, selected as the AWV input in the MOD/MYD11\_L2 retrieval, were extracted for comparison with the *in situ* AWV measured once a minute by an automatic Sun Sky Photometer CE-318. The absolute difference AWV values (ABS\_D\_AWV) between the *in situ* AWV values and MODIS AWV values from the corresponding pixels of MOD/MYD07\_L2 were calculated to show the relationship between the ABS\_D\_AWV and the absolute bias (ABS\_DT) of each MOD/MYD11\_L2 image in Fig. 6(a). Fig. 6(a) indicates that ABS\_DT is closely related with ABS\_D\_AWV, and their Pearson's  $r=0.87$ . Moreover, the ABS\_DT linearly increased with ABS\_D\_AWV to a certain extent, especially in the range of ABS\_D\_AWV > 0.4 cm, although there are too

few LST images validated in this study to obtain the rigid regression of their relationship. The distribution difference between the points of MOD11\_L2 (MOD, marked in black) and the points of MOD11\_L2 (MYD, marked in red) indicated that the ABS\_D\_AWV values of MOD07\_L2 are mostly greater than those of MYD07\_L2, and the average ABS\_D\_AWV of MOD07\_L2 (0.50 cm) is also larger than that of MYD07\_L2 (0.16 cm), which probably caused a relative lower accuracy of MOD11\_L2 in this validation case. Therefore, the error of input AWV could heavily increase the error of the MODIS LST products over a certain range, such as 0.4 cm in this validation case. Moreover, as indicated by Sobrino et al. (2014), the MODIS atmospheric profile, based on the knowledge of surface emissivities to separate the surface contribution from the sounding data, might not be very accurate in semi-arid and arid areas because of the high variation of surface emissivities. The increasing possibility of the low accurate MODIS AWV over the heterogeneous surface in semi-arid and arid areas probably become a considerable factor to reduce the accuracy of the MODIS LST.

Since Wan and Li (2008) have indicated that when the AWV is larger than 1.5 cm, the error of MODIS V5 L2 level daily LST gets larger as AWV increases, the relationship of the ABS\_DT and *in situ* AWV are also analyzed and shows in Fig. 6(b). Fig. 6(b) shows that there is no obvious similar trend between ABS\_DT and *in situ* AWV. However, they have a Pearson's  $r=0.50$ , and the point with a greater AWV probably has a larger ABS\_DT to some extent. Therefore, the refined GSW algorithm retrieving the Collection 6 MODIS 1 km daily LST products has a relatively greater tolerance to the large AWV than the old version GSW algorithm, which probably benefits from the wider range of the difference (from  $-10\text{ K}$  to  $29\text{ K}$ ) between LST and the air temperature at the surface level that has been set to regress the refined GSW coefficients (Wan, 2014).

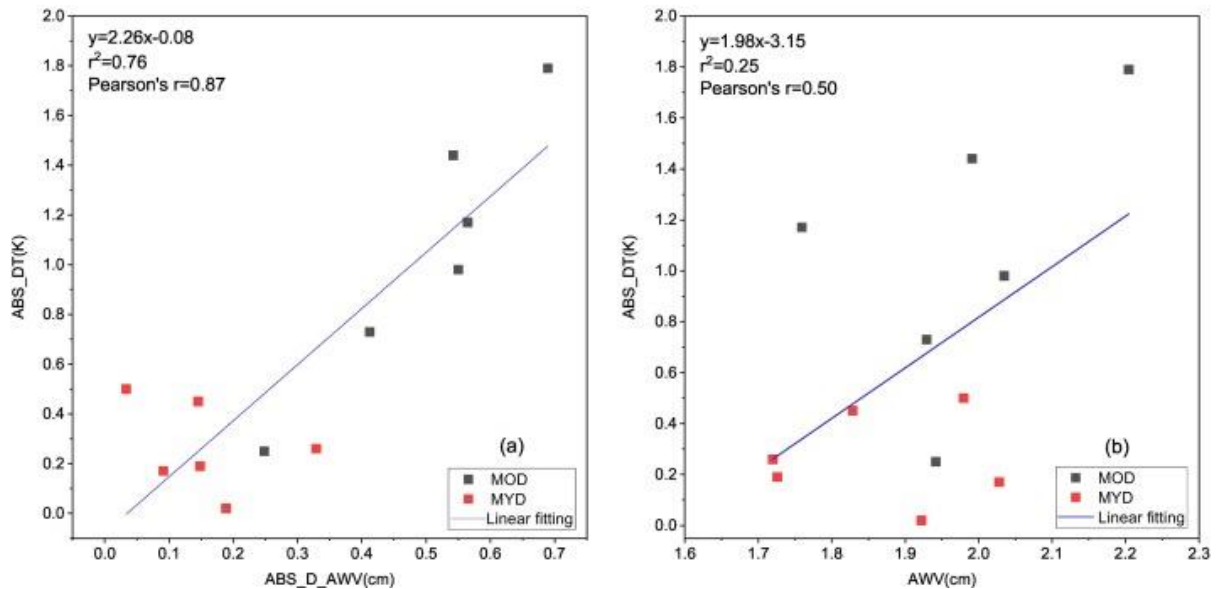


Figure 6. The impacts of the AWV on the error in the MODIS Level 2 LST products. (a) is the impact of the AWV uncertainty (absolute difference, ABS\_D\_AWV) of MODIS Collection 6 daily atmospheric profile products (MOD/MYD07) on absolute bias (ABS\_DT) of each MODIS LST image; and (b) is the impact of the AWV values on the absolute bias (ABS\_DT) of each MODIS LST image.

#### 4.2.3. Impacts of dust aerosols and AOD

The aerosol properties in TIR differ greatly due to complex and variable dust aerosols, so dust aerosols can influence the LSTs retrieved using the split-window algorithm (Wan and Dozier 1996) in two ways:  $T_{\text{mean}}$  and  $(T_{31} - T_{32})$ . The first change has been addressed and barely affects the accuracy of the MODIS Level 2 LSTs. Thus, the impacts of dust aerosols on LSTs may be reflected by changes in the term  $(T_{31} - T_{32})$ , which have been proven to have a large effect on the accuracy difference between MOD11\_L2 and MYD11\_L2 in Section 4.2.1. Therefore, to evaluate the impacts of dust aerosols on the term  $(T_{31} - T_{32})$ , the variation in the daytime atmospheric turbidity between the timing of the Terra and Aqua overpasses was selected to represent changes in dust aerosols.

Based on the Angstrom formula (equation (7)) indicating the dependence of the AOD on the wavelength (Wang et al. 2015), the atmospheric Angstrom turbidity coefficient was obtained from the AOD dataset (Wang et al., 2015):

$$\tau_{\lambda}^{\alpha} = \beta \lambda^{-\alpha} \quad (7)$$

where  $\tau_{\lambda}^{\alpha}$  is the AOD at  $\lambda$ ;  $\lambda$  is the wavelength measured in  $\mu\text{m}$ ;  $\alpha$  is the Angstrom exponent coefficient relevant to the size distribution of the aerosol particles; and  $\beta$  is the aerosol extinction coefficient when  $\lambda$  is 1  $\mu\text{m}$ .  $\beta$  is related to the amount of aerosol present in the atmosphere, also called the atmospheric turbidity or turbidity coefficient, and it is a dimensionless index of the opacity of a vertical column of the atmosphere.  $\beta$  is widely applied in meteorological environment detection and the measurement of air pollution. Therefore,  $\beta$  was selected in this study to represent daytime dust aerosols.

The field experiment was concurrent with both Terra and Aqua overpasses on 10<sup>th</sup> July 2012, so the daytime AOD data on this day, were selected to estimate the change in atmospheric turbidity ( $\beta$ ) in the daytime. Figure 7 shows that the atmospheric turbidity decreased slightly over time. However, the difference in the atmospheric turbidity at the overpass times of the Terra and Aqua satellites was very small and the difference between the maximum and minimum  $\beta$  on the selected day was less than 0.035.  $\beta$  of the AOD dataset selected in the ER-based validation process was compared with the Angstrom coefficient value under clear-sky conditions in North America (Grenier et al. 1995) and Europe (Boscà et al. 1996; Elminir et al. 2006). Most of the  $\beta$  values corresponding to the MODIS overpasses in this study are low, indicating that the ER-based process was applied under clear atmospheric conditions, and the aerosol concentrations were low and suitable for *in situ* calibration of remote sensing satellite sensors. Dust aerosols had only a weak influence on the difference in the ( $T_{31} - T_{32}$ ) values and on the effects in the range considered by the split-widow algorithm within the atmospheric condition simulation. Therefore, dust aerosol is not a main factor contributing to the error in MOD11\_L2.

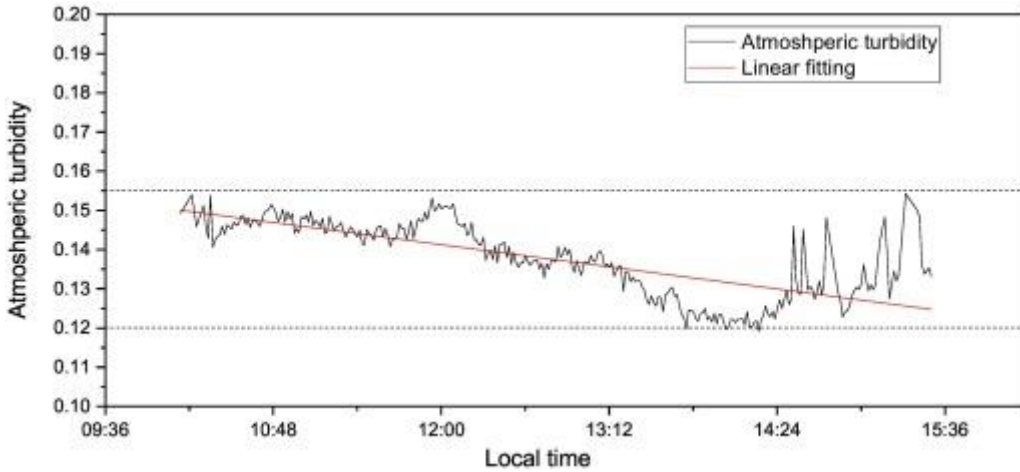


Figure 7. Change in the daytime Angstrom atmospheric turbidity.

To explore the effects of other atmospheric factors, mainly the AOD, the relationship between the mean errors of each image granule ID with AOD was analyzed (Fig. 8). To easily label the x-axis, the images were numbered in chronological order as case numbers. According to Fig. 8, the changes in the MODIS LST error ( $T_{\text{MODIS}} - T_{\text{in situ}}$ , namely, DT) is not sensitive to changes in AOD in bands 440 nm, 550 nm, 670 nm, 870 nm, and 1020 nm (Pearson  $r = 0.08$  and  $p = 0.786 > 0.05$ ), and there is an apparent relationship among the AODs in different bands, which can be explored further through Eq. (7).

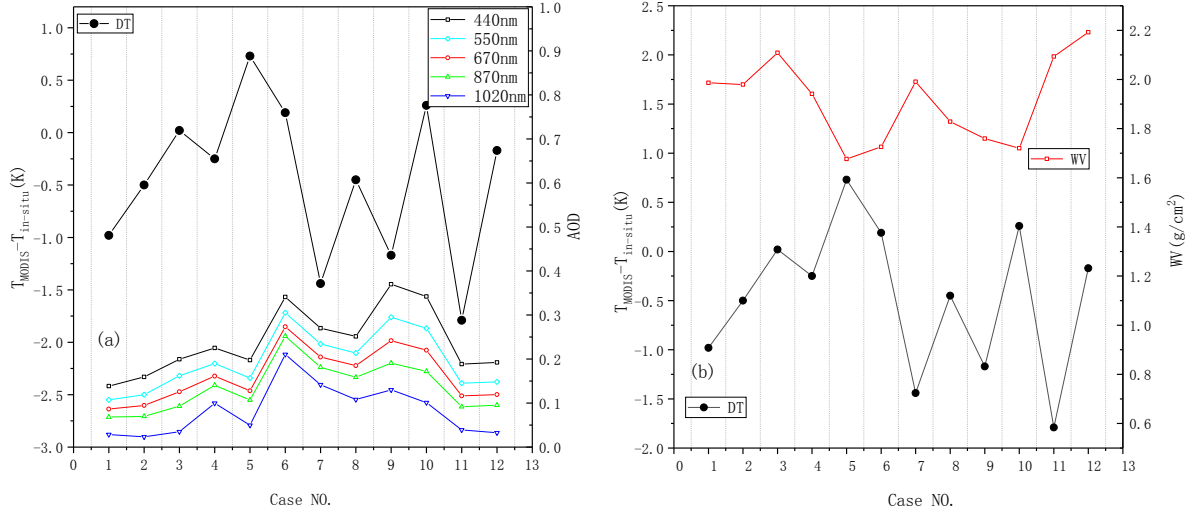


Figure 8. The effect of AOD and total AWW content

in the vertical path of the atmosphere on the errors in the MODIS Level 2 LST products.

#### 4.2.4. Effect of the mean brightness temperature ( $T_{mean}$ ) in bands 31 and 32

The  $T_{mean}$  in bands 31 and 32 corresponds to the term  $((T_i+T_j)/2)$  in Eq. (6). Since  $T_{mean}$  is an important input in the split-window algorithm, its effect on the errors in MOD11\_L2 and MYD11\_L2 are given in Fig. 9. Its potential contribution to the accuracy difference between the MOD11\_L2 and MYD11\_L2 products in the FEA is discussed below. Because the data points are randomly scattered in Fig. 9, there is no significant correlation between the errors of the daytime LSTs in both MOD11\_L2 and MYD11\_L2 and the corresponding  $T_{mean}$  values ( $|Pearson\ r| < 0.4$  and  $p > 0.01$ ). A comparison of Fig. 9(a) and Fig. 9(b) shows that there is no significant difference between the effect of the term  $((T_i+T_j)/2)$  on the MOD11\_L2 products and that on the MYD11\_L2 products. Thus, the term  $((T_i+T_j)/2)$  does not contribute to the large difference between the accuracies of MOD11\_L2 and MYD11\_L2.

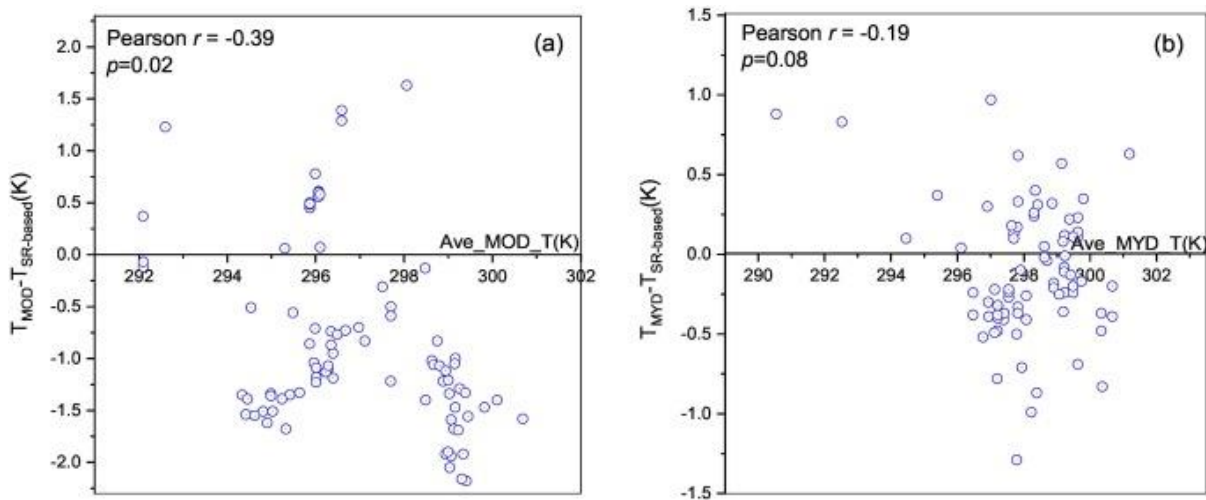


Fig. 9. Effects of the mean brightness temperature ( $T_{mean}$ ) in bands 31 and 32 on the errors in MOD11\_L2 and MYD11\_L2; (a) effect of  $T_{mean}$  (labeled Ave\_MOD\_T) on the error (labeled T<sub>MOD</sub>-T<sub>SR</sub>-based) in the MOD11\_L2 products, and (b) effect of  $T_{mean}$  (labeled Ave\_MYD\_T) on the error (labeled T<sub>MYD</sub>-T<sub>SR</sub>-based) in the MYD11\_L2 products.

### 4.3. Validation of MODIS emissivity products and their effect on the accuracy of MODIS LST retrievals

#### 4.3.1. Validation of the MOD11\_L2 and MYD11\_L2 products

To validate the emissivity involved in the MOD/MYD11\_L2 products, the temperature-independent band-average emissivity for a mixed pixel can be defined under the condition that other physical characteristics remain unchanged as the LST varies (Wan and Dozier, 1996):

$$\bar{\varepsilon}_i = \frac{\int_{\lambda_i(lower)}^{\lambda_i(upper)} \Psi(\lambda) (\sum_j^N p_j \varepsilon_j) d\lambda}{\int_{\lambda_i(lower)}^{\lambda_i(upper)} \Psi(\lambda) d\lambda} \quad (8)$$

where  $\lambda_i(upper)$  and  $\lambda_i(lower)$  are the upper and lower boundaries of band  $i$ , respectively;  $\Psi(\lambda)$  is the spectral response function of the sensor in band  $i$ ; and  $p_j$  is the proportion of LCT  $j$  in the mixed pixel. The temperature-independent band-average emissivity for mixed pixels (Wan and Dozier 1996) could be used to express the thermal emittance from mixed pixels in the split-window spectral region. Therefore, the *in situ* band-average emissivities of mixed pixels in MODIS bands 31 and 32 (in\_situ\_E31 and in\_situ\_E32) were calculated based on the ground-based effective spectral emissivities of mixed pixels from Equation (4). Table 3 shows an obvious overestimation in the MODIS band emissivity products compared to the *in situ* band-average emissivity, probably due to the underestimation of the proportion of bare land in the mixed pixels. According to a comparison of the LSE from MOD11\_L2 (data from the Terra platform) to those from MYD11\_L2 (data from the Aqua platform), the accuracy of the emissivity product from the two platforms is almost the same in average bias (AB) and RMSE. However, MOD\_E31 and MYD\_E31 have a larger error (AB of 0.019) than MOD\_32 and MYD\_E32 (AB of 0.012). Therefore, care should be taken when using the emissivity products from MOD11\_L2 and MYD11\_L2 directly without further comparison work, especially the emissivity products with heterogeneous surfaces.

Table 3 Emissivity comparison between MODIS MOD/MYD\_L2 and the *in situ* band-average emissivity. SI, AB, and RMSE are statistical indicator, average bias, and root-mean-square error. MOD\_E31 and MOD\_E32 are emissivity products from MOD11\_L2 for bands 31 and 32, and MYD\_E31 and MYD\_E32 are the emissivity data from MYD11\_L2.

SI	MOD_E31-in_situ_E31	MOD_E32-in_situ_E32	MYD_E31-in_situ_E31	MYD_E32-in_situ_E32
AB	0.0192	0.0118	0.0186	0.0120
RMSE	0.0193	0.0120	0.0189	0.0121

#### 4.3.2. Evaluation of the impacts of the emissivity validation on LST retrieval

According to previous study (Wan and Dozier 1996), the GSW LST algorithm is more sensitive to changes in the difference of LSEs between bands 31 and 32 than to changes in the LSEs themselves or the mean LSEs in bands 31 and 32. The average value ( $\Delta E_{MOD}$ ) of the difference between (MOD\_E31 - MOD\_E32) and (in\_situ\_E31 - in\_situ\_E32) and the average value ( $\Delta E_{MYD}$ ) of the difference between (MYD\_E31 - MYD\_E32) and (in\_situ\_E31 - in\_situ\_E32) were calculated to assess the effects of the emissivity on the accuracy of the MODIS LST retrieval using the split-window algorithm. The results show that  $\Delta E_{MOD}$  and  $\Delta E_{MYD}$  are 0.0072 and 0.0067, respectively; thus, the difference in the LSEs between bands 31 and 32 is overestimated by the MODIS emissivity data in MOD11\_L2 and MYD11\_L2. Because most of the difference values (E31-E32) are negative, the absolute difference in surface emissivity for bands 31 and 32 is smaller than that of the *in situ* band-average emissivity for bands 31 and 32. Therefore, it is reasonable to increase the absolute difference between bands 31 and 32 in the GSW algorithm to enhance its accuracy and stability in retrieving LSTs over mixed pixels. Despite the addition of a constraint for bare soil to adjust the difference between bands 31 and 32 during the new refinements in Collection 6, the mean values are unchanged for these pixels.

#### 4.4. Influence factors of the ER-based validation process

The MODTRAN radiation transfer model has been widely used in TIR-related studies, whereas the inputs are considered as the main factors influencing the SR-based method. The availability of accurate LCT data with a high spatial resolution is a precondition for the successful implementation of the SR-based method over a heterogeneous land surface. For a long-term or repeated validation, atmospheric profile data should be measured concurrently with the sensor overpass. Compared to the variation in the LST, the surface emissivity is relatively stable. Therefore, the effective spectral emissivity of the validated mixed pixels does not strictly require synchronous measurements with satellite observations. To calculate a reasonable effective spectral emissivity of the mixed pixels in the validation region, the FVC in an incompletely vegetated area must be obtained along with



the emissivity measurement as required supporting data. When the background aerosols are in excess of the average aerosol loadings, the AOD data obtained concurrently with the satellite data can be employed to eliminate the effects of dust aerosols on LSTs.

This study also validated the MODIS Collection 6 Level 2 LST data by the ER-based method. The LCT products with a 1 m spatial resolution collected via aerial remote sensing were used in the ER-based validation. Moreover, in the validation region, which was also selected as the investigation area for the LCT products, all vegetation types in the FEA LCT map were investigated in the field experiment. Therefore, the land cover data used in the ER-based validation can be considered as the “true” LCT data with a 1 m spatial resolution. The atmospheric profile was obtained by the sounding balloon (Table 1). From the launch information in the table (start time and end time of the launch of the sounding balloon), the MODIS overpass time occurred primarily during the local time period. Moreover, the launch site of the sounding balloons was in the middle of the 5 km × 5 km area on flat and open ground (Figure 1), and the sounding balloons were all launched under clear-sky conditions. Therefore, the atmospheric profiles used in the ER-based validation are representative atmospheric profile data from the FEA region and were collected concurrently with the MODIS overpasses. Although the atmospheric conditions of the ER-based validation are clear, the sun photometer observations in the FEA were used to provide the atmospheric aerosol content and AWV content for the simulated *in situ* LSTs. The effective spectral emissivity values of the element types in the FEA region were obtained almost synchronously with the experiment implemented to maintain the temporal representativeness of each endmember emissivity.

The LST error caused by the definition of the effective emissivity for the mixed pixels does not exceed 0.04 K, as indicated by Wan and Dozier (1996). To estimate the effect of the effective emissivity of a mixed pixel, the mixed pixels containing the largest number of endmembers were selected to simulate the LST errors added by the SR-based validation under low and high transmittance in band 31 of MODIS. A simple sensitivity analysis experiment was employed in this study with a series of biases from -0.01 to 0.01 (with a step of 0.005) added to the effective emissivity of the selected validated pixels in the MODIS LST products. The mixed pixels with the largest number of endmembers in the images with low transmittance (granule ID: A2012215.0415) and high transmittance (granule ID: A2012185.0405) were selected to estimate the effects on the MOD11\_L2 product. Similarly, the mixed pixels with the largest number of endmembers in the images with low transmittance (granule ID: A2012215.0555) and high transmittance (granule ID: A2012192.0550) were selected to estimate the effect on the MYD11\_L2 product. The other conditions were left unchanged, and the sensitivity of the accuracy of the emissivity measurements are shown in Fig. 10. The effective emissivity is a relatively sensitive input parameter for the SR-based validation process, and the absolute LST biases of the SR-based validation results affected by the error in the spectral emissivity were weakened by the low transmittance in band 31. In this study, the spectral emissivity was retrieved from the field-measured radiance observed by the portable 102F. Its lab test document reports that its spectral calibration is up to 1 wavenumber (Hook and Kahle, 1996). Under the calibration of warm and hot blackbodies coming with the instrument, the radiance error of the sample is <0.5 K. To ensure the accuracy of the spectral emissivity, the key steps suggested by Hook and Kahle (1996) for the field operation of 102F were followed, and the average emissivity spectrum of five measurements was taken as the emissivity spectrum for a given typical object. The spectral emissivity data were retrieved via the LECTES approach and the validated results indicate that the mean error of LSE is <0.01 (Wang et al., 2011). Therefore, with a bias of 0.01 added for each endmember in each wavelength region, the error that the spectral emissivity adds to the SR-based process is far <0.7 K. Moreover, the error decreases in the upscaling process from 102F observations to MODIS observations and the averaging process of endmembers emissivities in the mixed pixel.

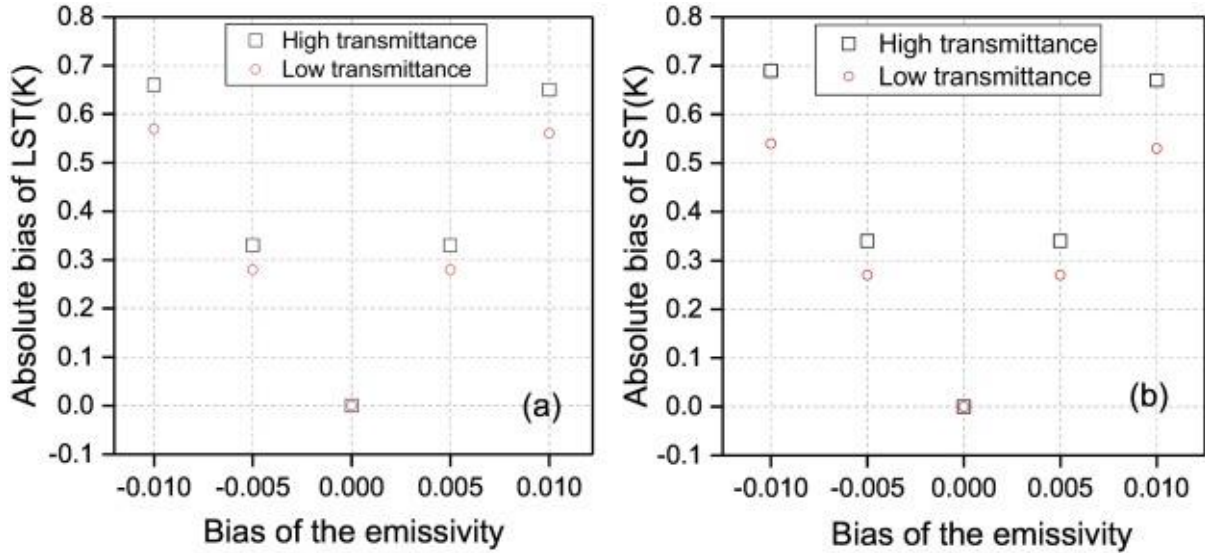


Figure 10. Sensitivity analysis of the emissivity measurements; (a) the emissivity accuracy effect on the validation of the MOD11\_L2 product using the ER-based method, and (b) the emissivity accuracy effect on the validation of the MYD11\_L2 product using the ER-based method.

## 5. Conclusions

This paper presents a SR-based approach for validating satellite-derived LST products over heterogeneous land surfaces, which are widely observed worldwide. This approach supplements the R-based method by incorporating the effective spectral emissivity. Compared to the T-based validation method, the greatest advantage of the SR-based method is no requirement of a ground-based LST which are difficult to obtain over heterogeneous pixels. The SR-based validation approach is also used to validate Collection 6 Level 2 daily LST/LSE products (MOD11\_L2 and MYD11\_L2) which have not been adequately validated since their release. The validation work in the FEA indicates that the daytime MOD11\_L2 and MYD11\_L2 products have accuracies of  $-0.84 \pm 0.88$  K and  $-0.11 \pm 0.42$  K, respectively. The analysis of the possible factors contributing to the difference between the MOD11\_L2 and MYD11\_L2 products reveals the significant relation between  $(T_{31} - T_{32})$  and MOD11\_L2 errors, which is primarily caused by the AWW accuracy difference between MODIS atmospheric profile products from Terra (MOD07\_L2) and Aqua (MYD07\_L2) as AWW input to the LST retrieval algorithm.

Additionally, the LSE products in bands 31 and 32 are validated by the *in situ* band-average emissivity. The results show that there is an obvious overestimation in the MODIS band emissivity products over heterogeneous land surface, probably because of the underestimation of the proportion of bare land in the mixed pixels. Since E31-E32 is a sensitive term in the split-widow algorithm, the effect of emissivity validation results on the LST retrieval can be evaluated based on the average difference ( $\Delta E$ ) between E31-E32 in the MODIS LSE products and E31-E32 *in situ* band-average emissivity. The results show that  $\Delta E$  of MOD11\_L2 and MYD11\_L2 are 0.0072 and 0.0067, respectively; thus, E31-E32 is overestimated by the MODIS emissivity data in the MOD11\_L2 and MYD11\_L2 products. Moreover, the absolute MODIS E31-E32 values are underestimated because the E31-E32 values are mostly negative in both the MODIS emissivity and *in situ* data. Therefore, increase in the absolute emissivity difference between bands 31 and 32 in the split-widow algorithm is probably necessary to improve the LST accuracy in mixed pixels in arid regions, such as the HRB region.

An accurate RTM is the basis of the SR-based method. Ground-based measurements and aerial remote sensing data that are synchronous or quasisynchronous with MODIS overpasses are crucial to implementing the SR-based method over heterogeneous land surfaces. The requisite data mainly include atmospheric profiles, spectral emissivity values, and land cover images with a high spatial resolution. The accuracy of the SR-based results is dependent on the quality of the input data. All these datasets may limit the use of this method in long-term validation. However, compared to other currently acceptable validation methods for heterogeneous surfaces, such as systematic SSN validation strategies that require multisampling strategies and scale transfer technologies, the SR-based method is an effective and relatively simple method for validating satellite-derived LST products, especially LST products with a relatively low spatial resolution in areas where it is difficult to obtain pixel-scale LST measurements.

Similar as most studies, there are limitations in our research. Firstly, the successful SR-based validation process depends strictly on the accuracy of concurrent *in situ* measurements. Moreover, the results in this study may be limited by the small experiment region and the limited frequency. Therefore, further validation of the

MODIS Collection 6 Level 2 daily LST/LSE products should be implemented widely to provide for local or even global accuracy references of mixed pixels for their application. Our method was developed and tested in the HRB, Northwest China. However, the SR-based process is a general validation framework and can be widely applied in other areas and for other aerospace-based LST products, as long as the local concurrent *in situ* measurements can be collected.

### **Acknowledgments**

This study was supported by the Natural Science Foundation of China (Grant No. 41601448 and 41830648) and the Fundamental Research Funds for the Central Universities of China (Grant No. XDJK2019C016). The authors thank all of the groups that contributed to the HiWATER experiment for supporting the original measurements and releasing the data online for free. All the MODIS data in this study were acquired from the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Archive Center (DAAC), located in the Goddard Space Flight Greenbelt, Maryland (<https://ladsweb.nascom.nasa.gov/>). The authors are also grateful to Dr. Zheng from ICube at the University of Strasbourg for his helpful suggestions.

### **Author Contributions**

Wenping Yu outlined the research, was responsible with the data preprocessing and analysis, and wrote the manuscript. Mingguo Ma and Hong Yang provided helpful suggestions on the research design and helped revise the manuscript. Junlei Tan and Xiaolu Li collected and preprocessed all data involved in this study in the data.

### **References**

- Anderson, M.C., Allen, R.G., Morse, A., & Kustas, W.P., 2012. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. *Remote Sensing of Environment*, 122, 50-65.
- Arnfield, A.J., 2003. Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*, 23, 1-26.
- Bastiaanssen, W.G.M., Noordman, E.J.M., Pelgrum, H., Davids, G., Thoreson, B.P., & Allen, R.G., 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering-Asce*, 131, 85-93.
- Berk, A., Anderson, G.P., Acharya, P.K., Fox, M., Adlergolden, S.M., Hoke, M.L., Gardner, J.A., Cooley, T.W., Borel, C.C., & Lewis, P.E., 2005. MODTRAN 5: a reformulated atmospheric band model with auxiliary species and practical multiple scattering options: update. *Proceedings of Spie the International Society for Optical Engineering*, 5806, 662-667.
- Boscà, J.V., Pinazo, J.M., Cañada, J., & Ruiz, V., 1996. Ångström's turbidity coefficient in Seville, Spain in the years 1990 and 1991. *International Journal of Ambient Energy*, 17, 171-178.
- Brunsell, N.A., & Gillies, R.R., 2003. Length Scale Analysis of Surface Energy Fluxes Derived from Remote Sensing. *Journal of Hydrometeorology*, 4, 1212-1219.
- Cheng, G.D. (2009). *Integrated Management of the Water-Ecology-Economy system in Heihe River Basin*. Beijing, China: Science Press
- Coll, C., Caselles, V., Galve, J.M., Valor, E., Niclos, R., Sanchez, J.M., & Rivas, R., 2005. Ground measurements for the validation of land surface temperatures derived from AATSR and MODIS data. *Remote Sensing of Environment*, 97, 288-300.
- Coll, C., Caselles, V., Valor, E., & Niclòs, R., 2012a. Comparison between different sources of atmospheric profiles for land surface temperature retrieval from single channel thermal infrared data. *Remote Sensing of Environment*, 117, 199-210.

- Coll, C., Valor, E., Galve, J.M., Mira, M., Bisquert, M., García-Santos, V., Caselles, E., & Caselles, V., 2012b. Long-term accuracy assessment of land surface temperatures derived from the Advanced Along-Track Scanning Radiometer. *Remote Sensing of Environment*, 116, 211-225.
- Coll, C., Wan, Z., & Galve, J.M., 2009. Temperature-based and radiance-based validations of the V5 MODIS land surface temperature product. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 114
- Duan, S.-B., Li, Z.-L., Wu, H., Leng, P., Gao, M., Wang C., 2018. Radiance-based validation of land surface temperature products derived from Collection 6 MODIS thermal infrared data. *Int. J. Appl. Earth Obs. Geoinf.*, 70 84-92.
- Elminir, H.K., Hamid, R.H., El-Hussainy, F., Ghitas, A.E., Beheary, M.M., & Abdel-Moneim, K.M., 2006. The relative influence of the anthropogenic air pollutants on the atmospheric turbidity factors measured at an urban monitoring station. *Science of The Total Environment*, 368, 732-743.
- Ermida, S.L., Trigo, I.F., DaCamara, C.C., Göttsche, F.M., Olesen, F.S., & Hulley, G., 2014. Validation of remotely sensed surface temperature over an oak woodland landscape — The problem of viewing and illumination geometries. *Remote Sensing of Environment*, 148, 16-27.
- Grenier, J.C., Casinière, A.D.L., & Cabot, T., 1995. Atmospheric Turbidity Analyzed by Means of Standardized Linke's Turbidity Factor. *Journal of Applied Meteorology*, 34, 1449-1458.
- Hook, S.J., & Kahle, A.B., 1996. The micro Fourier transform interferometer ( $\mu$ FTIR)—a new field spectrometer for acquisition of infrared data of natural surfaces. *Remote Sensing of Environment*, 56, 172-181.
- Hoover, G., & Kahle, A.B., 1987. A thermal emission spectrometer for field use. *Photogrammetric Engineering and Remote Sensing*, 53, 627-632.
- Huang, C., Chen, W., Li, Y., Shen, H., & Li, X., 2016. Assimilating multi-source data into land surface model to simultaneously improve estimations of soil moisture, soil temperature, and surface turbulent fluxes in irrigated fields. *Agricultural and Forest Meteorology*, 230-231, 142-156.
- Kogan, F.N., 2001. Operational Space Technology for Global Vegetation Assessment. *Bulletin of the American Meteorological Society*, 82, 1949-1964.
- Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., Qi, Y., Wen, J., Li, H., Zhu, G., Guo, J., Ran, Y., Wang, S., Zhu, Z., Zhou, J., Hu, X., & Xu, Z., 2013a. Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific Objectives and Experimental Design. *Bulletin of the American Meteorological Society*, 94, 1145-1160.
- Li, X., Li, X., Li, Z., Ma, M., Wang, J., Xiao, Q., Liu, Q., Che, T., Chen, E., Yan, G., Hu, Z., Zhang, L., Chu, R., Su, P., Liu, Q., Liu, S., Wang, J., Niu, Z., Chen, Y., Jin, R., Wang, W., Ran, Y., Xin, X., & Ren, H., 2009. Watershed Allied Telemetry Experimental Research. *Journal of Geophysical Research*, 114.
- Li, X., Shaomin, L., Xiao, Q., Mingguo, M., Jin, R., Che, T., Wang, W., Hu, X., Xu, Z., Wen, J., & Wang, L., 2017. A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system. *Scientific Data*, 4, 170083.
- Li, Z.-L., Tang, B.-H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., & Sobrino, J.A., 2013b. Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, 131, 14-37.

- Martin, M.A., Ghent, D., Pires, A.C., Göttsche, F.-M., Cermak, J., Remedios, J.J. 2019. Comprehensive in situ validation of five satellite land surface temperature data sets over multiple stations and years. *Remote Sens.*, 11, 479.
- McMillin, L.M., 1975. Estimation of sea surface temperatures from two infrared window measurements with different absorption. *Journal of Geophysical Research*, 80, 5113-5117.
- Pinheiro, A.C.T., Privette, J.L., & Guillevic, P., 2006. Modeling the observed angular anisotropy of land surface temperature in a Savanna. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1036-1047.
- Pinker, R.T., Sun, D., Hung, M.-P., Li, C., & Basara, J.B., 2009. Evaluation of satellite estimates of land surface temperature from GOES over the United States. *Journal of Applied Meteorology and Climatology*, 48, 167-180.
- Prata, F., 2002. Land Surface Temperature Measurement from Space: AATSR Algorithm Theoretical Basis Document. *Contract Report to Esa*, 508-510.
- Rozenstein, O., Qin, Z., Derimian, Y., & Karnieli, A., 2014. Derivation of Land Surface Temperature for Landsat-8 TIRS Using a Split Window Algorithm. *Sensors*, 14, 5768-5780.
- Salomonson, V.V., Barnes, W., Maymon, P.W., Montgomery, H.E., & Ostrow, H., 1989. MODIS: advanced facility instrument for studies of the Earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, 27, 145-153.
- Slater, P.N., Biggar, S.F., Thome, K.J., Gellman, D.I., & Spyak, P.R., 1996. Vicarious radiometric calibrations of EOS sensors. *Journal of Atmospheric and Oceanic Technology*, 13, 349-359.
- Smith, P.M., Kalluri, S.N.V., Prince, S., & Defries, R., 1997. The NOAA/NASA Pathfinder AVHRR 8-km land data set. *Photogrammetric Engineering and Remote Sensing*, 63(1), 12-27.
- Snyder, W.C., Wan, Z., Zhang, Y., & Feng, Y.Z., 1998. Classification-based emissivity for land surface temperature measurement from space. *International Journal of Remote Sensing*, 19, 2753-2774.
- Sobrino, J., Jimenez-Munoz, J.-C., Mattar, C., Sòria Barres, G. 2014. Evaluation of Terra/MODIS atmospheric profiles product (MOD07) over the Iberian Peninsula: a comparison with radiosonde stations. <http://repositorio.uchile.cl/handle/2250/138320>
- Trigo, I.F., Monteiro, I.T., Olesen, F., & Kabsch, E., 2008a. An assessment of remotely sensed land surface temperature. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 113
- Trigo, I.F., Peres, L.F., DaCamara, C.C., & Freitas, S.C., 2008b. Thermal land surface emissivity retrieved from SEVIRI/Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 307-315.
- Wan, Z., 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, 140, 36-45.
- Wan, Z.M., Dozier, J. 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Trans. Geosci. Remote Sens.*, 34, 892-905.
- Wan, Z., & Li, Z.-L., 1997. A physics-based algorithm for retrieving land-surface emissivity and temperature from EOS/MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 980-996.
- Wan, Z., & Li, Z.L., 2008. Radiance-based validation of the V5 MODIS land-surface temperature product. *International Journal of Remote Sensing*, 29, 5373-5395.

- Wan, Z., Zhang, Y., Zhang, Q., & Li, Z.-l., 2002. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment*, 83, 163-180.
- Wan, Z.M., & Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *Ieee Transactions on Geoscience and Remote Sensing*, 34, 892-905.
- Wang, L., Gong, W., Xia, X., Zhu, J., Li, J., & Zhu, Z., 2015. Long-term observations of aerosol optical properties at Wuhan, an urban site in Central China. *Atmospheric Environment*, 101, 94-102.
- Wang, N., Wu, H., Nerry, F., Li, C., & Li, Z.-L., 2011. Temperature and Emissivity Retrievals From Hyperspectral Thermal Infrared Data Using Linear Spectral Emissivity Constraint. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 1291-1303.
- Wang, S., Li, X., Ge, Y., Jin, R., Ma, M., Liu, Q., Wen, J., & Liu, S., 2016. Validation of Regional-Scale Remote Sensing Products in China: From Site to Network. *Remote Sensing*, 8, 980.
- Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 335-344.
- Yu, W.P., Ma, M.G., Wang, X.F., Song, Y., & Tan, J.L., 2011. Validation of MODIS land surface products using ground measurements in the Heihe River Basin. *Proceedings of SPIE*, 8714, 817423-817422.
- Yu, W., Ma, M., Li, Z., Tan, J., Wu, A. 2017. New scheme for validating remote-sensing land surface temperature products with station observations. *Remote Sens.*, 9, 1210.
- Zhang, X., Zhou, J., Göttsche, F.-M., Zhan, W., Shaomin, L., Cao, R. 2019. A method based on temporal component decomposition for estimating 1-km all-weather land surface temperature by merging satellite thermal infrared and passive microwave observations. *IEEE Trans. Geosci. Remote Sens.* ,1-21
- Zhao, S., Cong, D., He, K., Yang, H., Qin, Z. 2017. Spatial-temporal variation of drought in China from 1982 to 2010 based on a modified Temperature Vegetation Drought Index (mTVDI). *Sci. Rep.*, 7.